ESCUELA DE DOCTORADO
INTERNACIONAL DE LA USC

Cristina
Ameneiro Quiñoy

Tesis doctoral

Estudio de la función CCM en pluripotencia y especificación endotelial

Santiago de Compostela, 2023

Programa de doctorado en Medicina Molecular

TESIS DE DOCTORADO

ESTUDIO DE LA FUNCIÓN CCM EN PLURIPOTENCIA Y ESPECIFICACIÓN ENDOTELIAL

Cristina Ameneiro Quiñoy

ESCUELA DE DOCTORADO INTERNACIONAL DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOCTORADO EN MEDICINA MOLECULAR

SANTIAGO DE COMPOSTELA

2023

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

DECLARACIÓN DEL AUTOR/A DE LA TESIS

D./Dña. **Cristina Ameneiro Quiñoy**

Título da tese: **Estudio de la función CCM en pluripotencia y especificación endotelial**

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento y declaro que:

1) La tesis abarca los resultados de la elaboración de mi trabajo.
2) De ser el caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.
3) Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.
4) La tesis es la versión definitiva presentada para su defensa y coincide la versión impresa con la presentada en formato electrónico.

Y me comprometo a presentar el Compromiso Documental de Supervisión en el caso que el original no esté depositado en la Escuela.

En **Santiago de Compostela**, **15 de diciembre de 2022**.

Firma electrónica

AUTORIZACIÓN DEL DIRECTOR/TUTOR DE LA TESIS

D./Dña. **Miguel Ángel Fidalgo Pérez**

En condición de: **Tutor/a y director/a**

Título de la tesis: **Estudio de la función CCM en pluripotencia y especificación endotelial**

INFORMA:

Que la presente tesis, se corresponde con el trabajo realizado por D/Dña **Cristina Ameneiro Quiñoy**, bajo mi dirección/tutorización, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director/tutor de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

En **Santiago de Compostela**, **15 de diciembre de 2022**

Firma electrónica

AUTORIZACIÓN DEL DIRECTOR/TUTOR DE LA TESIS

D./Dña. **Diana Guallar Artal**

En condición de: **Director/a**

Título de la tesis: **Estudio de la función CCM en pluripotencia y especificación endotelial**

INFORMA:

Que la presente tesis, se corresponde con el trabajo realizado por D/Dña **Cristina Ameneiro Quiñoy**, bajo mi dirección/tutorización, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director/tutor de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

En **Santiago de Compostela**, **15 de diciembre de 2022**

Firma electrónica

DECLARACIÓN DE CONFLICTO DE INTERESES Y AUTORÍA DE LOS DATOS PRESENTADOS

Cristina Ameneiro Quiñoy declara no tener ningún conflicto de interés en relación con la presente tesis doctoral.

Por la presente, la doctoranda declara la completa autoría de todas las imágenes y figuras mostradas en la tesis.

En Santiago de Compostela, 20 de marzo de 2023

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

La presente tesis doctoral se realizó gracias a las “ayudas de apoyo a la etapa predoctoral en las universidades del SUG, en los organismos de investigación de Galicia y en otras entidades del Sistema gallego de I+D+i, cofinanciadas parcialmente por el programa operativo FSE Galicia 2014-2020, para el 2020” (Referencia de la ayuda: ED481A-2020/046)

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

AGRADECIMIENTOS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Llega ese ansiado y temido momento. El momento en el que se materializa el final de una etapa. Echando la vista atrás, se suceden infinidad de recuerdos que permanecerán siempre conmigo, pues tuve la gran suerte de que este tiempo me aportó mucho más de lo que habría esperado nunca, tanto a nivel científico como personal. Por eso quiero recordar y agradecer a todas las personas que se involucraron y me ayudaron a llegar hasta aquí.

En primer lugar, agradecer a mis directores de tesis, el doctor Miguel Fidalgo y la doctora Diana Guallar, por confiar en mí y acogerme como estudiante de máster, dándome la oportunidad de participar en diversos proyectos de investigación a lo largo de este tiempo. A ellos les debo casi todo lo que aprendí no solo de pluripotencia y epigenética, sino de cómo enfrentarme a los desafíos que acompañan al mundo de la ciencia. Gracias por vuestra pasión y dedicación, y por saber ser mis mentores.

A todos los miembros de la unidad Stem Cells & Epitranscriptomics, a los veteranos y a los que se fueron uniendo a esta pequeña familia, en especial a Vera y Adriana. Sin vosotras no hubiese sido lo mismo. Gracias por el apoyo, las charlas, los cotilleos, las risas en Ámsterdam y las salidas a comer.

A mis amigas Carmen y Patri. Por acompañarme todos estos años y crecer conmigo, sois parte de quien soy hoy. Gracias por contenerme en los mejores y peores momentos, aunque yo no lo pusiera siempre fácil.

A Paula, la mejor hermana del mundo, y más importante, mi amiga. Siempre has sido y serás un ejemplo para mí. Gracias por aguantarme, escucharme, aconsejarme, ayudarme y cuidarme siempre. A Gabriel, gracias por contagiarnos alegría y felicidad desde el primer día. Con tu inocencia, cariño e ingenio consigues que los problemas sean menos problema. A Ale, me resulta imposible recordar lo que significa no tenerte en mi vida, más que mi cuñado, mi hermano. Gracias por preocuparte siempre.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A mis padres, Miguel y Rosa, gracias por todo lo que me disteis y por todo lo que me enseñasteis. Vuestro apoyo fue lo que me trajo hasta aquí, por animarme siempre a perseguir lo que me hace feliz. Mi padre me enseñó el significado de resiliencia y de ser incondicional, de él aprendí a seguir siempre adelante. De mi madre aprendí lo que es la bondad y la entrega desinteresada, siempre serás la mejor persona que conoceré, y siempre te llevaré conmigo e intentaré hacer justicia a tus valores. Os quiero.

Por último, y especialmente, a Álex. Mi mejor amigo, mi compañero, mi persona en el mundo. Gracias por estar a mi lado y hacer este camino más fácil, supiste convertir todos los días malos en días felices. Me animas, me apoyas y me inspiras a dar lo mejor de mí en todos los aspectos. Llenaste estos años de momentos mágicos. Te quiero.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A Álex
A mi familia
A ti

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

TABLA DE CONTENIDO

1. **RESUMEN** 5
2. **INTRODUCCIÓN** 9
2.1 El estado celular pluripotente 11
2.1.1 Desarrollo embrionario temprano 11
2.1.2 Células troncales embrionarias 13
2.1.3 Estados alternativos de pluripotencia 14
2.2 Mecanismos moleculares de regulación de la pluripotencia 18
2.2.1 Control transcripcional 18
2.2.2 Control epigenético 20
2.2.2.1. Metilación del ADN 20
2.2.2.2. Modificaciones de histonas 23
2.2.3 Señalización celular 26
2.2.3.1. Vías de señalización en estados alternativos de pluripotencia 27
2.2.3.2. Vía de señalización FGF-ERK1/2 32
2.3 Salida del estado celular pluripotente 34
2.4 Proteínas CCM 36
2.4.1 Estructura y función de las proteínas CCM 36
2.4.2 Papel de las CCM en desarrollo embrionario 41
2.4.3 CCMs en enfermedad 42
3. **OBJETIVOS** 47
4. **MATERIAL Y MÉTODOS** 51
4.1 Reactivos químicos y Kits 53
4.2 Modelos experimentales 53
4.3 Cultivo celular 54
4.4 Clonación de vectores plasmídicos 57
4.5 Transformación y obtención de vectores plasmídicos 59
4.6 Transfección de células de mamífero 61
4.7 Generación de partículas lentivirales 62

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.8 Transducción de células de mamífero 63
4.9 Generación de líneas *knock-out* mediante CRISPR-Cas9 63
4.10 Extracción y purificación de ADN genómico 64
4.11 Extracción y purificación de ARN 65
4.12 Obtención de ADN complementario 65
4.13 Análisis de expresión génica mediante qPCR o PCR 66
4.14 Extracción y purificación de proteína 67
4.15 Análisis mediante Western blot 69
4.16 Tinción de la fosfatasa alcalina 70
4.17 Inmunofluorescencia 70
4.18 Análisis de ciclo celular 71
4.19 Citometría de flujo 72
4.20 Ensayos de proliferación 72
4.21 Ensayos de formación de colonias pluripotentes 73
4.22 Análisis de los niveles de *Nanog*-GFP 73
4.23 Inmunoprecipitación 74
4.24 Determinación de la metilación genómica 75
4.25 Ensayos de diferenciación 76
4.26 Ensayo de especificación endotelial 77
4.27 Secuenciación masiva de ARN y análisis 78
4.28 Análisis *in silico* de los datos transcriptómicos 79
4.29 Inmunoprecipitación de cromatina seguida de secuenciación masiva 82
4.30 Análisis de perfiles de enriquecimiento a partir de ChIP-seq 87
4.31 Análisis de las proteínas que interactúan con CCM2 87
4.32 Análisis de redes de interacción 90
4.33 Análisis estadísticos 90
4.34 Recursos bibliográficos 90
4.35 Software y algoritmos 91
5. **RESULTADOS** 93
5.1 Analizar el papel de los genes *Ccm* en el mantenimiento de pluripotencia 95
5.1.1 Los genes *Ccm* muestran patrones de expresión específicos 95
5.1.2 Generación de ESCs KO para cada gen *Ccm* 97
5.1.3 La función CCM es dispensable en el mantenimiento de pluripotencia 99

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.1.4 La ausencia de CCM2 favorece la autorrenovación de las ESCs 102
5.1.5 Las CCMs intervienen en el paisaje transcripcional pluripotente 105
5.1.6 La ausencia de CCM2 promueve pluripotencia *naïve* 111
5.1.7 CCM2 contribuye a la salida de la pluripotencia *naïve* 114
5.1.8 CCM2 es necesario para establecer y mantener la pluripotencia *primed*. 117
5.2 Elucidar los programas moleculares gobernados por CCM2 en pluripotencia .. 122
5.2.1 CCM2 no interviene en el control de la expresión génica a nivel de cromatina 122
5.2.2 CCM2 es necesario para la activación de ERK1/2 126
5.2.3 CCM2 modula indirectamente la actividad de ERK1/2 131
5.2.4 La inhibición de ERK1/2 contribuye a la desregulación transcripcional causada por la pérdida de CCM2 135
5.2.5 PRC2 participa en los cambios transcripcionales tras la pérdida de CCM2 140
5.2.6 Las *Ccm2* KO presentan niveles reducidos de metilación en ADN 148
5.2.7 La pérdida de CCM2 altera la abundancia de enzimas que gobiernan la metilación genómica 149
5.2.8 DNMTs y TETs intervienen en el paisaje transcripcional de las *Ccm2* KO . 151
5.2.9 Identificación de los interactores de CCM2 156
5.2.10 Predicción de interactores de CCM2 envueltos en la regulación de pluripotencia *naïve* 160
5.2.11 CCM2 vincula múltiples vías epigenéticas a la red de pluripotencia 163
5.3 Determinar la función CCM en diferenciación y especificación endotelial 166
5.3.1 La función CCM participa en la regulación de programas transcripcionales de diferenciación celular 166
5.3.2 Las ESCs deficientes en CCMs muestran una diferenciación anormal hacia cuerpos embrioides 169
5.3.3 La eliminación de las CCM compromete la gastrulación en cultivo 171
5.3.4 Optimización y validación de la diferenciación hacia células progenitoras endoteliales a partir de ESCs 173
5.3.5 Las proteínas CCM son necesarias para generar EPCs 178
5.3.6 La pérdida de cada CCM altera genes comunes relacionados con la función endotelial 181
5.3.7 Identificación de intervenciones farmacológicas potenciales para modular la ausencia de función CCM en el contexto de especificación endotelial 185

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6. **DISCUSIÓN** 189
6.1 Proteínas CCM en enfermedad 194
6.2 Proteínas CCM en desarrollo 197
6.3 Mecanismo molecular de la señalización CCM 199
6.4 Moduladores de la señal CCM 202
7. **CONCLUSIONES** 205
8. **APÉNDICES** 209
8.1 Tabla suplementaria 1. Reactivos y químicos 211
8.2 Tabla suplementaria 2. Kits 214
8.3 Tabla suplementaria 3. Accesorios y fungible 215
8.4 Tabla suplementaria 4. Oligonucleótidos 217
8.5 Tabla suplementaria 5. Anticuerpos 219
8.6 Tabla suplementaria 6. Datasets 222
8.7 Tabla suplementaria 7. Ontologías 223
8.8 Tabla suplementaria 8. Lista de interactores de CCM2 225
8.9 Tabla suplementaria 9. Lista con los principales genes desregulados en *Ccms* KO identificados por RNA-seq 228
9. **LISTA DE ABREVIATURAS** 241
10. **LISTA DE FIGURAS Y TABLAS** 249
11. **BIBLIOGRAFÍA** 257

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. RESUMEN

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Las células pluripotentes, debido a sus propiedades de autorrenovación ilimitada en cultivo y capacidad de dar lugar a cualquier tipo celular de un individuo adulto, poseen un gran potencial en medicina regenerativa y representan un poderoso sistema para estudiar los fundamentos moleculares que controlan el destino celular. Mientras que el control de pluripotencia se encuentra conservado en mamíferos, todavía no se dispone de un conocimiento completo de los fundamentos moleculares que gobiernan el estado celular pluripotente, lo que limita sus fines terapéuticos. Aunque la pérdida de función completa de uno de los tres genes *Ccm* (*Ccm1, Ccm2* o *Ccm3*) conlleva alteraciones en la función endotelial, que resulta en malformaciones cavernosas cerebrales o en letalidad embrionaria en ratones, todavía se desconoce el papel de cada proteína CCM en pluripotencia y en la toma de decisiones celulares que determinan el linaje endotelial. En base a esto, el objetivo principal de mi tesis doctoral consistió en descubrir la función de cada proteína CCM en los complejos mecanismos moleculares que gobiernan la pluripotencia y la especificación endotelial. Para ello, utilicé como modelo principal de estudio células trocales embrionarias (ESCs) y realicé una amplia variedad de ensayos celulares y bioquímicos, técnicas ómicas y análisis bioinformáticos. Como resultado, identifiqué que la función CCM es importante para el control transcripcional que gobierna la pluripotencia y descubrí que su ausencia altera el proceso correcto de diferenciación celular. De manera destacada, encontré que CCM2 coordina estados alternativos de pluripotencia a través de la vía de señalización ERK1/2 y de mecanismos epigenéticos. Por otro lado, determiné una firma de expresión génica que interviene en la especificación endotelial, lo que me permitió demostrar que la función de cada proteína CCM es necesaria para generar células progenitoras endoteliales a partir de ESCs. En conjunto, mis resultados proporcionan un conocimiento más profundo de la función CCM en los programas moleculares que controlan la pluripotencia y que permiten una correcta diferenciación celular. Además, los mecanismos identificados abren nuevas vías de estudio para entender mejor la función CCM en desarrollo temprano, biología endotelial y enfermedad.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2. INTRODUCCIÓN

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A lo largo de las siguientes secciones, con el fin de facilitar la comprensión de los resultados obtenidos durante mi tesis doctoral, se abordará de manera simple y concisa el estado del arte relacionado con la pluripotencia y mecanismos moleculares asociados, así como el papel de las CCMs en biología endotelial, desarrollo, y enfermedad. Me centraré en pluripotencia murina por ser el modelo de estudio escogido para interrogar qué funciones desempeñan las proteínas CCM en pluripotencia y especificación del linaje endotelial.

2.1 EL ESTADO CELULAR PLURIPOTENTE

El potencial de diferenciación de una célula viene determinado por su plasticidad y su capacidad para dar lugar a otros tipos celulares más especializados. Así, dependiendo de su potencial de diferenciación, nos encontramos ante un espectro que comprende células totipotentes (capaces de dar lugar tanto a tejidos embrionarios como extraembrionarios), pluripotentes, multipotentes, oligopotentes y células unipotentes, en orden decreciente de potencial de diferenciación[1]. El estado celular pluripotente, que existe únicamente durante el desarrollo temprano embrionario, se encuentra definido por el potencial para diferenciar en células pertenecientes a las capas germinales (ectodermo, mesodermo y endodermo) y a las células germinales primordiales, que en conjunto darán lugar a todos los tejidos que conformarán un organismo adulto[1].

2.1.1 Desarrollo embrionario temprano

El desarrollo embrionario murino comienza con dos células especializadas provenientes de una hembra y un macho, oocito y espermatozoide respectivamente, que se fusionan para dar lugar a un zigoto. Tras este momento, en el cual se establece la totipotencia, el potencial de diferenciación se irá restringiendo a medida que continúa el proceso de desarrollo (**Figura I1**). Tras la fertilización, el desarrollo del zigoto se encuentra controlado por factores maternos presentes en el oocito[2], críticos para la activación del genoma embrionario al

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

posibilitar la expresión de nuevos transcritos necesarios para continuar el desarrollo[3]. Entre los días de desarrollo embrionario E1 y E2.5, el zigoto progresa hacia los estados de 2-células, 4-células y 8-células, para a continuación, transitar por un proceso de polarización, compactación, y cavitación de la mórula, hasta la formación del blastocisto temprano a día E3.5[4-6]. En este punto, se produce la primera especificación de linaje, estando el blastocisto conformado por una capa externa de células denominada trofectodermo (TE) y una masa celular interna (ICM, del inglés *inner cell mass*)[7-11]. Mientras que las células del TE formarán los componentes fetales de la placenta, las células de la ICM constituyen el conjunto de células pluripotentes que finalmente darán lugar a un organismo completo. Las células de la ICM expresan genes específicos como *Oct4*, *Nanog*, *Sox2* y *Esrrb*[12], todos ellos factores reguladores críticos del estado celular pluripotente como se mencionará en detalle más adelante. Seguidamente, estas células sufren una segunda segregación, dando lugar a las células del epiblasto temprano y del endodermo primitivo a día E4.5[11]. El epiblasto da lugar al propio feto, mientras que el endodermo primitivo se convierte en el endodermo parietal y visceral, contribuyendo eventualmente al saco vitelino[13]. Esta segunda especificación aparece determinada, entre otros eventos, por la vía de los receptores del factor de crecimiento fibroblástico (FGFR, del inglés *fibroblast growth factor receptor*), y se encuentra mediada por la quinasa proteína-quinasa activada por mitógenos MEK (del inglés *mitogen-activated protein kinase kinase*), que interviene en la activación de la quinasa regulada por señal extracelular ERK1/2 (del inglés *extracelular signal regulated kinase*)[11,14]. Al igual que las células de la ICM, las células del epiblasto temprano mantienen un estado celular pluripotente. A continuación, el blastocisto se implanta en la pared endometrial uterina, un proceso intrincado que necesita que todos los pasos se produzcan con éxito: control hormonal para que el útero sea receptivo, liberación del blastocisto de su capa externa (zona pelúcida), adhesión del blastocisto al endometrio, e invasión de las células del trofoblasto unido a la reconstrucción arterial materna para asegurar el flujo sanguíneo al

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

embrión[15]. A día E5.5, el epiblasto maduro ya se ha implantado en el endometrio y comienza la gastrulación, proceso que inicia la especificación de las capas germinales[16]. Durante esta ventana del desarrollo embrionario temprano murino, que constituye los días pre-implantación E3.5-E4.5 y el período post-implantación E5.5-E6.5, las células de la ICM y del epiblasto mantienen la capacidad pluripotente.

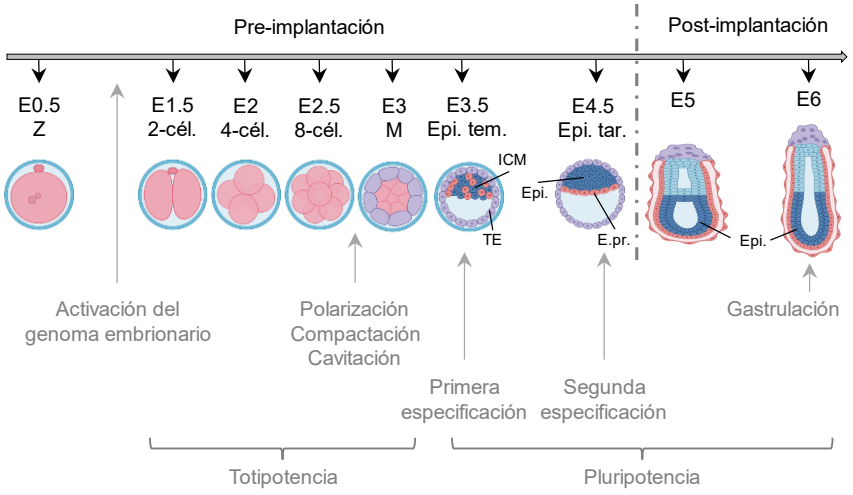


Figura I1. Desarrollo embrionario temprano. Representación esquemática simplificada de las etapas tempranas del desarrollo embrionario murino. Se muestra mediante flechas el momento embrionario en el que ocurren los principales procesos pre- y post-implantación. Se especifica el potencial de diferenciación asociado a cada etapa (totipotencia o pluripotencia), representándose en azul oscuro las células que se encuentran en un estado pluripotente (ICM y epiblasto). Z: zigoto, 2-cél.: embrión de 2 células, 4-cél.: embrión de 4 células, 8-cél.: embrión de 8 células, M: mórula, Epi. tem.: epiblasto temprano, Epi. tar.: epiblasto tardío, ICM: masa celular interna, TE: trofectodermo, Epi.: epiblasto, E. pr.: endodermo primitivo. Imagen creada con BioRender.com.

2.1.2 Células troncales embrionarias

El potencial embrionario se puede mantener en cultivo mediante la derivación de células de embriones E3.5-E6.0, dando lugar a diferentes tipos de células pluripotentes cuyas características dependen

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

del momento del desarrollo en el que se encuentra el embrión del que son extraídas[17-21]. Al contrario que las células pluripotentes involucradas en el desarrollo embrionario, que solo existen durante una pequeña ventana temporal, las células derivadas *ex vivo* pueden mantenerse indefinidamente bajo condiciones de cultivo que favorecen su autorrenovación[22,23].

Las primeras líneas pluripotentes murinas aisladas, denominadas células troncales embrionarias (ESCs, del inglés *embryonic stem cells*), se derivaron a partir de las células de la ICM de blastocistos de pre-implantación a día E3.5[24,25]. Posteriormente, se aislaron células troncales de epiblasto (EpiSCs, del inglés *epiblast stem cells*), derivadas del epiblasto de blastocistos de post-implantación a días E5.5 y E5.75[18,20]. Así, las ESCs y las EpiSCs representan una herramienta extraordinaria para estudiar y entender tanto los mecanismos que controlan y regulan el estado celular pluripotente, como la salida de éste durante el desarrollo, con el resultante comienzo de los procesos de diferenciación. Estos mecanismos están sujetos a una exhaustiva regulación del control transcripcional, epigenético y de diferentes vías de señalización celular.

2.1.3 Estados alternativos de pluripotencia

Dependiendo del momento del desarrollo, las células derivadas del blastocisto presentarán unos estados celulares y moleculares característicos, definidos como *naïve*, intermediarios y *primed* (**Figura I2**). Dichos estados son distinguibles a través de la morfología de las colonias celulares, el requerimiento de distintos factores para su crecimiento, el metabolismo energético, la firma molecular y el estado de inactivación del cromosoma X en el caso de las células femeninas[17,26,27]. Las características principales que describiré a continuación se encuentran resumidas en la **Tabla I1**.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

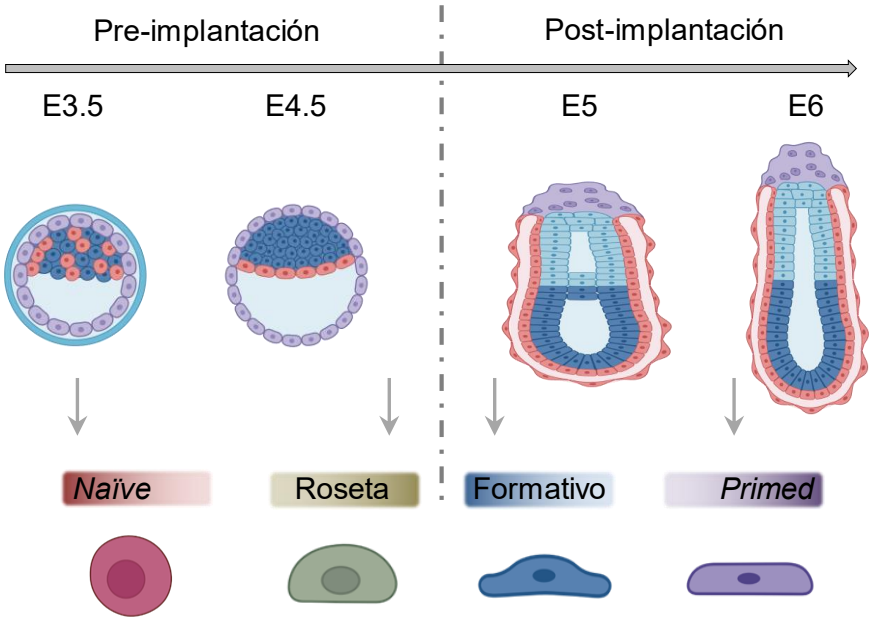


Figura I2. Derivación *ex vivo* de células pluripotentes. Representación esquemática simplificada de las etapas tempranas de desarrollo embrionario y los correspondientes estados alternativos pluripotentes derivados *ex vivo*. Imagen creada con BioRender.com.

Las células pluripotentes *naïve* mimetizan a las células de la ICM del blastocisto temprano a día E3.5[17]. En cultivo, estas células crecen en colonias pequeñas, compactas y redondas. Las células *naïve* son capaces de formar quimeras cuando se inyectan en embriones pre-implantacionales, al incorporarse tanto en las tres capas germinales como en las células del linaje germinal primordial (PGC, del inglés *primordial germ cell*), aunque con menos eficiencia en este último caso[28]. Además, estas células presentan las dos copias del cromosoma X activas[17]. Para conservar este estado alternativo de pluripotencia en cultivo, las células se mantienen en ausencia de suero, y en presencia del factor inhibidor de la leucemia (LIF, del inglés *leukaemia inhibitory factor*), un inhibidor de la glucógeno sintasa quinasa-3 beta, denominada GSK3β (CHIR99021), y un inhibidor de MEK

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(PD0325901)[17,29]. GSK3β actúa como regulador negativo de la vía de señalización de WNT[30,31], mientras que MEK es un componente integral de la vía de los factores de crecimiento fibroblástico (FGF) y ERK1/2[32].

Cuando las ESCs derivadas de la ICM del blastocisto temprano a día E3.5 se mantienen bajo condiciones pluripotentes estándar en cultivo (esto es en presencia de suero y LIF) su estado pluripotente celular se convierte en heterogéneo, recibiendo el nombre de metaestable. Esta heterogeneidad implica la coexistencia de células en un estado más próximo al *naïve* y células más avanzadas en el espectro pluripotente, viéndose afectada tanto la morfología de las colonias como la expresión de marcadores pluripotentes y modificaciones epigenéticas[33,34].

Recientemente se ha descrito un estado intermediario pluripotente adicional, denominado en cultivo como células troncales en forma de roseta[35]. Estas células son similares a las células del epiblasto justo antes de la implantación del blastocisto, cuando se vuelven polarizadas[36]. En cultivo, estas células pluripotentes crecen en colonias más dispersas, incluso llegando a verse planas en algunos casos. Al igual que las células pluripotentes *naïve*, tienen la capacidad de formar quimeras, y presentan activas las dos copias del cromosoma X[35]. En este caso, el estado celular pluripotente se puede mantener en cultivo en presencia de LIF, un inhibidor de la vía de señalización de WNT (IWP2) e inhibición de MEK con PD0325901[35].

Otro estado pluripotente intermediario es el denominado formativo, análogo a las células del epiblasto recién implantado[37]. Estas células presentan una morfología de colonia plana. Al igual que los estados pluripotentes anteriores, las células del estado formativo son capaces de contribuir a la formación de quimeras, y además son capaces de dar lugar de manera eficiente al linaje germinal primordial[19]. Por otro lado, una de las copias del cromosoma X ya se encuentra parcialmente inactivada[37]. Estas células se mantienen en cultivo en ausencia de suero

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

y en presencia de una baja concentración de la citoquina ACTIVINA A, un inhibidor de la vía de señalización de WNT (XAV939) y un inhibidor de la vía del ácido retinoico (BMS493)[19].

El estado pluripotente más restringido y preparado para diferenciarse es el *primed*, que mimetiza las células del epiblasto justo antes de la gastrulación[18,20], siendo denominadas células troncales del epiblasto (EpiSCs, del inglés *epiblast stem cells*) las que se derivan directamente de estas etapas del desarrollo murino. Las colonias formadas por estas células presentan una morfología plana, creciendo en monocapa[20]. Las células pluripotentes *primed* pierden la capacidad de contribuir a la formación de quimeras[20,28,38] y la segunda copia del cromosoma X se encuentra completamente inactivada[37]. En cultivo, se mantienen en ausencia de suero y en presencia de altas concentraciones de ACTIVINA A y FGF2[38,39].

Tabla I1. Comparación de los distintos estados alternativos pluripotentes

Característica	*Naïve*	Roseta	Formativo	*Primed*
Morfología	Redondas	Redondeadas	Aplanadas	Planas
Equivalente embrionario	E3.5-4.5	E5	E5.5-6.0	E6.5-7.0
Contribución a quimeras	Alta	Alta	Baja	Baja-nula
Contribución a PGC	Baja	Baja	Alta	Muy baja
Medio de cultivo	LIF, iMEK, iGSK3β	LIF, iWNT, iMEK	↓ActA, iWNT, iRA	↑ActA, FGF2
Proliferación cél. única[40]	Sí	Sí	Sí	No
Estado del cromosoma X	XaXa	XaXa	XaXa-i	XaXi
Metilación de ADN[37,40]	Baja	Media	Media	Alta

Xa: cromosoma X activo, Xi: cromosoma X inactivo, ActA: activina A, RA: ácido retinoico, del inglés *retinoic acid*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.2 MECANISMOS MOLECULARES DE REGULACIÓN DE LA PLURIPOTENCIA

El establecimiento y mantenimiento del estado pluripotente depende de una red génica de regulación y de un estricto control de la interacción y cooperación de diversos factores de transcripción y reguladores epigenéticos. Debido a la poca información que existe aún acerca del estado pluripotente intermedio representado por las células troncales en forma de roseta, centraré este apartado en los estados *naïve*, formativo y *primed*, abarcando así el espectro pluripotente descrito.

2.2.1 Control transcripcional

La red génica de regulación de la pluripotencia es un sistema altamente interconectado, donde los factores de transcripción OCT4, NANOG y SOX2 representan el eje central, presente en todos los estados pluripotentes. Estos tres factores actúan en sincronía, regulando sus propios promotores génicos y conformando un circuito autoregulatorio[41-44] (**Figura I3**). Esta red de proteínas regula a otros factores secundarios con el fin tanto de mantener el estado pluripotente como de reprimir la señalización y expresión de factores de diferenciación.

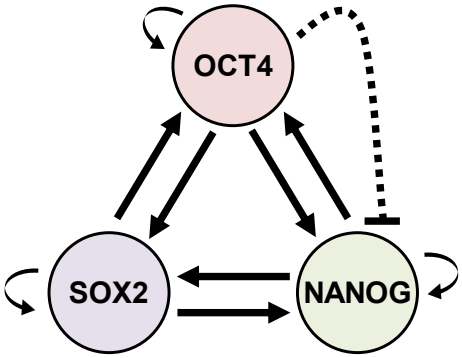


Figura I3. Circuito autorregulatorio de los factores centrales en el control transcripcional del estado pluripotente. OCT4, SOX2 y NANOG forman un circuito interconectado al actuar como factores de transcripción tanto sobre sus propios

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

promotores como entre ellos. De esta forma, se mantiene la expresión de los tres factores conservando un estado pluripotente estable.

Las células pluripotentes *naïve* expresan los tres factores que conforman el eje central, pero se distinguen de los otros estados pluripotentes por su mayor expresión de *Nanog*[17,45]. Además, también presentan niveles elevados en la expresión de factores de transcripción característicos como *Rex1*, *Klf4*, *Klf5*, *Esrrb*, *Dppa3*, *Tfcp2l1*, *Tbx3* y *Prdm14*[22,46,47]. Curiosamente, la expresión de *Oct4* se regula a través de dos regiones potenciadoras distintas, una proximal y otra distal, de las cuales las células *naïve* emplean preferencialmente su potenciador distal, al contrario que las formativas y las *primed*[40,48].

Las células intermediarias formativas sufren una completa reorganización del transcriptoma, desmantelándose parcialmente la red reguladora anterior[19]. En el estado formativo gana importancia el factor OTX2, cuya función principal consiste en redistribuir OCT4 hacia otros elementos reguladores del genoma[49,50]. Además, en este estado el gen *Oct4* pasa de regularse a través de su potenciador distal a hacerlo preferencialmente a través de su potenciador proximal[40,48]. A diferencia del estado *naïve* y en menor medida que el *primed*, las células formativas muestran una reducción en los niveles de expresión de *Nanog*[37,51]. Por otro lado, aumenta la expresión de marcadores como *Abcg2*, *Cldn4*, *Vgll1*, *Gata2*, *Gata3* y *Erp27*[51].

El estado pluripotente *primed*, al igual que los anteriores, también presenta expresión del eje central pluripotente *Oct4*, *Nanog* y *Sox2*. Estas células se caracterizan por un descenso en la expresión de los factores *naïve*[18], y un aumento de expresión de marcadores post-implantacionales determinantes de las futuras capas germinales de mesodermo (p. ej. *Brachyury*, *Eomes*, *Gsc*, *Mixl1* y *Fgf8*), endodermo (p. ej. *Sox17*, *Gata6*, *Gata4* y *FoxA2*) y ectodermo (p. ej. *Oct6*, *Nodal*, *Fgf5* y *Lefty*)[20].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.2.2 Control epigenético

El término "epigenética", acuñado por Conrad Hall Waddington, define el conglomerado de mecanismos y fenómenos que pueden producir cambios en el fenotipo de una célula sin afectar su secuencia de ADN, e incluyen las modificaciones químicas del ADN (p. ej. metilación) y las modificaciones químicas en las colas de las histonas. Para mantener el estado pluripotente, así como para promover cambios de identidad celular, las modificaciones epigenéticas juegan un papel crítico[52]. Por ejemplo, el paso de células *naïve* a células formativas o *primed* requiere una controlada remodelación epigenética[19,53-55].

2.2.2.1. Metilación del ADN

La metilación del ADN es una modificación química covalente que consiste en la adición de un grupo metilo en el quinto carbono de los nucleósidos citosina (5-metilcitosina o 5-mC). La abundancia y distribución de las citosinas metiladas en el genoma se correlaciona con el estado transcripcional y con la expresión espaciotemporal génica característica de distintos tipos celulares. La metilación suele estar implicada en la represión transcripcional, y es esencial para el desarrollo y la susceptibilidad a enfermedades en mamíferos[56]. El ciclo de la metilación del ADN (**Figura I4**) se divide en tres fases: el establecimiento de la metilación *de novo*, el mantenimiento de la metilación durante la replicación y la desmetilación[57-59].

Las enzimas principales que se encargan de establecer los patrones específicos de metilación son las ADN metiltransferasas 3 alfa y 3 beta (DNMT3A/DNMT3B, del inglés *de novo DNA methyltransferases*)[57]. Además, existe una *de novo* metiltransferasa inactiva catalíticamente, DNMT3L, que es capaz de interactuar y modular la actividad tanto de DNMT3A como de DNMT3B[60]. El

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

mantenimiento de los patrones existentes de metilación es llevado a cabo por DNMT1 en conjunto con la proteína UHRF1[58,61].

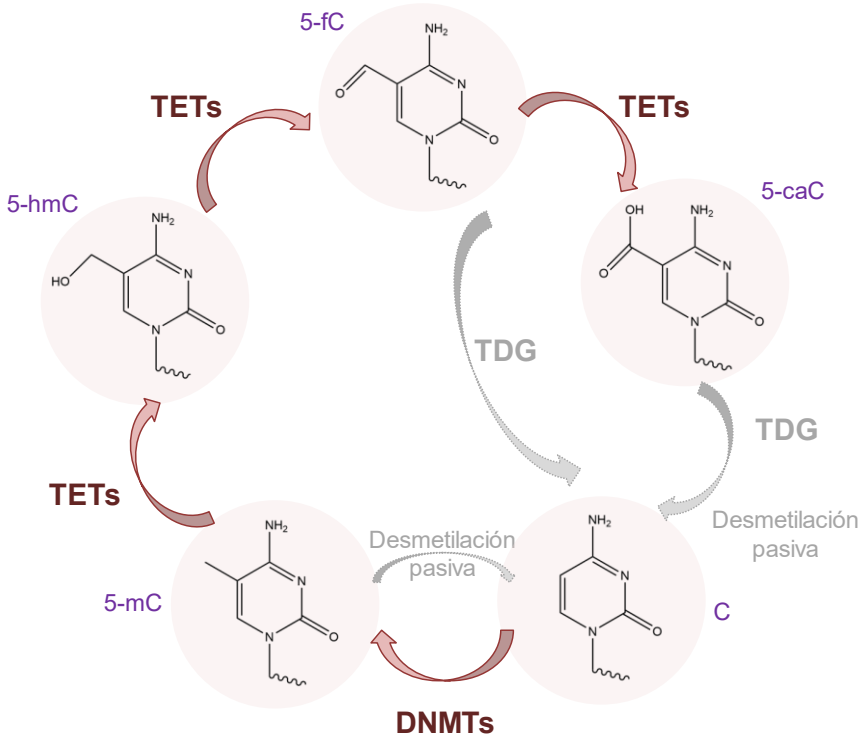


Figura I4. Ciclo de la metilación y desmetilación activa del ADN en células de mamífero. Las enzimas DNMT (del inglés *DNA methyltransferases*) depositan y mantienen grupos metilo en citosinas específicas (5-mC). Las enzimas TET (del inglés *ten-eleven translocation*) realizan sucesivas oxidaciones sobre las citosinas metiladas: 5-metilcitosinas a 5-hidroximetilcitosinas (5-hmC), 5-formilcitosinas (5-fC) y 5-carboxilcitosinas (5-caC), para contrarrestar el efecto de 5-mC. La enzima TDG es capaz de reconocer y eliminar las citosinas con modificaciones formil y carboxil iniciando la vía de reparación por escisión de bases, reemplazándose finalmente por citosinas sin modificar. Todas las modificaciones pueden perderse por un mecanismo pasivo a través de la dilución durante la replicación del ADN.

La desmetilación del ADN puede llevarse a cabo de forma pasiva, mediante la inhibición de las DNMT durante la replicación celular, o de forma activa. La desmetilación activa la llevan a cabo las

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

enzimas metilcitosina dioxigenasas TET (del inglés *ten-eleven translocation*), que oxidan iterativamente la 5-metilcitosina a 5-hidroximetilcitosina (5-hmC), 5-formilcitosina (5-fC) y 5-carboxilcitosina (5-caC)[62-64] (**Figura I4**). Además, todas las formas oxidadas promueven la desmetilación durante la replicación celular[65,66]. Por otro lado, la ADN glucosilasa de timina (TDG, del inglés *thymine DNA glycosylase*), puede eliminar directamente las bases 5-fC y 5-caC, entrando en juego la vía de reparación por escisión de bases[59,67] (**Figura I4**). Se han descrito tres enzimas TET: TET1, TET2 y TET3. TET3 se encuentra prácticamente restringida al oocito y al zigoto, mientras que TET1 y TET2 incrementan su expresión en el blastocisto y también se expresan en diversos tipos celulares del individuo adulto[68].

El estado pluripotente *naïve* muestra hipometilación global del ADN, constituyendo esta disminución de los niveles de 5-mC y de la expresión de *Dnmt3a* y *Dnmt3b* uno de los principales sellos intrínsecos de este estado[69,70]. Por el contrario, las células pluripotentes cultivadas en presencia de suero, que como hemos visto anteriormente representan un estado pluripotente heterogéneo, muestran un ligero incremento de los niveles globales de metilación del ADN y de expresión de las *de novo* metiltransferasas[71,72]. Con respecto a la desmetilación de las citosinas, nos encontramos con estudios opuestos, que describen bien un incremento o una invariabilidad en los niveles de 5-hmC en el estado *naïve*[69,73], bien una disminución de esta marca epigenética[74]. El estado *naïve* se caracteriza por un incremento global en los niveles proteicos de TET2[73], así como una disminución de TET1 a nivel global[69] o a nivel de su abundancia en la cromatina[73].

Los estados pluripotentes formativo y *primed* presentan un incremento gradual en los niveles globales de metilación del ADN, siendo el estado *primed* el que muestra mayores niveles, en correlación con una mayor expresión de *Dnmt3a* y *Dnmt3b*[73,75-77]. Además, presentan un aumento en los niveles de 5-hmC[76], y una disminución de TET2 en comparación con el estado *naïve*[73].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.2.2.2. Modificaciones de histonas

Las modificaciones de histonas son modificaciones covalentes postraduccionales que se dan predominantemente en las colas amino-terminales[78,79]. Entre el conjunto de enzimas modificadoras de histonas, las más estudiadas son aquellas que modifican su estado de metilación y de acetilación. Como para esta tesis son especialmente relevantes los complejos de *Polycomb* y la familia de metiltransferasas de histonas MLL, centraré este apartado en contextualizar estos reguladores epigenéticos en pluripotencia.

Los complejos represores 1 y 2 de *Polycomb* (PRC1 y PRC2, del inglés *Polycomb repressor complex*) se han descrito históricamente como represores epigenéticos[80], que son esenciales para un desarrollo embrionario correcto[81,82]. PRC1 deposita un grupo ubiquitina en la lisina 119 de la histona H2A (H2AK119ub)[83] mientras que PRC2 es el encargado de metilar la lisina 27 de la histona H3 (H3K27me3)[84]. PRC1 se subdivide en dos complejos principales, el canónico (cPRC1) y el no canónico (ncPRC1)[85] (**Figura I5**). Estos complejos presentan tanto funciones individuales como solapadas[86]. Por su parte, PRC2 también comprende dos complejos, denominados PRC2.1 y PRC2.2[87] (**Figura I5**). PRC1 y PRC2 poseen la capacidad de unirse a las marcas H3K27me3 y H2AK119ub, respectivamente, reforzando su actividad mutua[88]. Atendiendo a la función represiva clásica de los complejos de *Polycomb*, en el estado pluripotente PRC1 y PRC2 se encargan de regular genes involucrados en los procesos de desarrollo para así mantener la pluripotencia[81]. Posteriormente, se han descrito funciones no clásicas, asociadas a genes activos transcripcionalmente durante desarrollo y diferenciación celular[89-92]. Estos datos sugieren que la composición de las subunidades que conforman los complejos de *Polycomb* dictan si actúan como activadores o represores.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

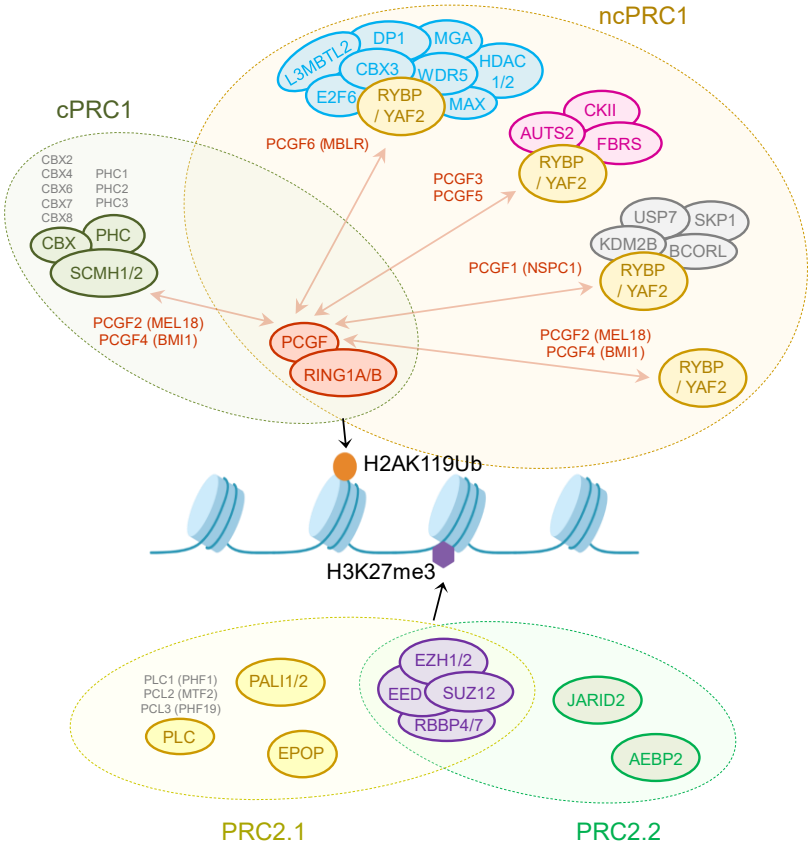


Figura I5. Composición de los complejos proteicos *Polycomb*. Los complejos represores de *Polycomb* (PRC) constituyen reguladores epigenéticos de la transcripción génica. Principalmente se forman dos complejos denominados PRC1 (que deposita un grupo ubiquitina en la lisina 119 de la histona H2A) y PRC2 (que deposita un grupo metilo en la lisina 27 de la histona H3). Ambos complejos constan de un núcleo que comprende una subunidad catalítica junto con cofactores que son esenciales para su actividad enzimática. A su vez, PRC1 puede conformar un complejo canónico (cPRC1), caracterizado por la interacción de la proteína RING con PCGF2 o 4 junto con las proteínas CBX, y un complejo no canónico (ncPRC1) que comprende RYBP o su homólogo YAF, y puede diferenciarse aún más debido a la subunidad de PCGF que se una. PRC2 se divide en PRC2.1 y PRC2.2 dependiendo de los cofactores que se incorporen. Imagen creada con BioRender.com.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

La familia de metiltransferasas de histonas MLL (del inglés, *mixed lineage leukemia*) es la responsable de la metilación de la lisina 4 de la histona 3, marca que se considera activadora en regiones potenciadoras (H3K4me1) y promotoras (H3K4me3)[93]. Esta familia comprende las proteínas MLL1/2/3/4 (KMT2A, KMT2B, KMT2C y KMT2D respectivamente) y SETDB1A/B, y es esencial durante el desarrollo embrionario[94,95]. Estas proteínas actúan junto con otros componentes básicos de los complejos MLL, como WDR5, RBBP5 y ASH2L[96,97].

Existe un grupo de genes de desarrollo cuyos promotores se encuentran marcados tanto con la marca de activación H3K4me3 como con la marca de represión H3K27me3, denominándose así genes bivalentes[98]. La bivalencia es una característica peculiar de los estados pluripotentes post-implantacionales, especialmente el estado *primed*, y está ausente tanto en el estado *naïve* como en las células diferenciadas[98]. En estos genes, estas dos marcas se dan en el mismo octámero, pero en histonas opuestas[99,100], y MLL2 es el encargado de depositar la marca H3K4me3[98,101,102].

El estado pluripotente *naïve* muestra un enriquecimiento global de H3K27me3, así como de los componentes centrales del complejo PRC2 (EZH2 y SUZ12) y varios componentes secundarios[103]. A su vez, estas células presentan incrementados los niveles de varios componentes de PRC1[103]. A pesar de ello, cuando nos centramos en genes de desarrollo, el estado pluripotente *naïve* presenta una reducción de la marca H3K27me3 y una limitada presencia de genes bivalentes[104,105]. La deleción del componente EED de PRC2 en este estado pluripotente resulta en una ganancia de los niveles de metilación del ADN, lo que sugiere una dependencia de PRC2 para mantener el estado hipometilado característico del estado pluripotente *naïve*[103]. Mientras que EZH2, SUZ12 y EED no son necesarios para mantener la autorrenovación ilimitada de las células pluripotentes, RING1A/B son imprescindibles para prevenir la diferenciación prematura en cultivo[87,106,107]. Por otro lado, aunque las proteínas MLL no son

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

necesarias para la autorrenovación de las ESCs, son imprescindibles para un correcto proceso de diferenciación celular[108-110].

Respecto al estado formativo, encontramos un aumento tanto de la marca activadora H3K4me3 como de la represora H3K27me3 en los promotores bivalentes que, aunque se encuentran listos para expresarse, no están activos (*poised*)[103,111]. Un número significativo de los genes bivalentes codifican reguladores de diferenciación, y no se encuentran activos en células pluripotentes debido a la presencia de una variante hipofosforilada de la polimerasa II de ARN, que los mantiene en este estado *poised*[105]. Durante el proceso de diferenciación, los genes bivalentes pierden una de las dos marcas, y con ello se define si su expresión es activada o permanece reprimida, dependiendo de la trayectoria celular final[112].

2.2.3 Señalización celular

El control del estado pluripotente mediante vías de señalización se lleva a cabo por el contacto entre células durante el desarrollo. Existen proteínas específicas secretadas que interaccionan con los receptores de la superficie celular, activando cascadas y factores de transcripción que regulan la expresión génica con el fin de controlar el destino celular, incluyendo el mantenimiento del estado celular pluripotente y los procesos de diferenciación. *Ex vivo*, la regulación de los distintos estados pluripotentes se lleva a cabo con la adición al medio de cultivo de moduladores que activan (activadores) o inhiben (inhibidores) las vías de señalización que rigen el momento análogo del desarrollo embrionario (**Figura I6**). Las vías más importantes se detallan a continuación.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

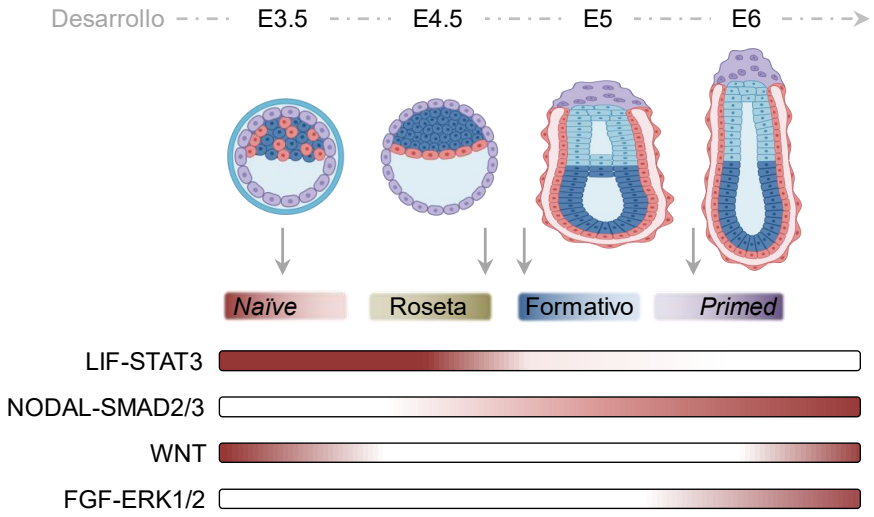


Figura I6. Activación de las vías que rigen el espectro pluripotente. Representación esquemática simplificada de las etapas del blastocisto peri-implantación mostrando el estado de activación de las vías LIF-STAT3, NODAL-SMAD2/3, WNT y FGF-ERK1/2 en el espectro pluripotente. El grado de intensidad del color granate representa el nivel de activación. Imagen creada con BioRender.com.

2.2.3.1. Vías de señalización en estados alternativos de pluripotencia

El factor inhibidor de la leucemia (LIF) es un miembro de la familia de citoquinas interleucina-6[113]. Este ligando se une a su receptor de la superficie celular (LIFR), activando la vía de señalización LIF/STAT3. La unión de LIF provoca la heterodimerización del receptor con la glicoproteína transductora de la señal 130 (GP130). Esta unión permite la activación de la quinasa asociada JAK, que a su vez fosforila LIFR y GP130, facilitando así la unión de STAT3 al dominio citoplasmático de los receptores[114] (**Figura I7**). Una vez unido STAT3 al receptor, es fosforilado por JAK, lo que permite su liberación, homodimerización y translocación al núcleo, donde actúa como factor de transcripción regulando la expresión de genes diana como *Klf4*,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Sox2, *Oct4*, *Nanog* y *Tfcp2l1*, favoreciendo así el mantenimiento del estado pluripotente en ESCs[115-118].

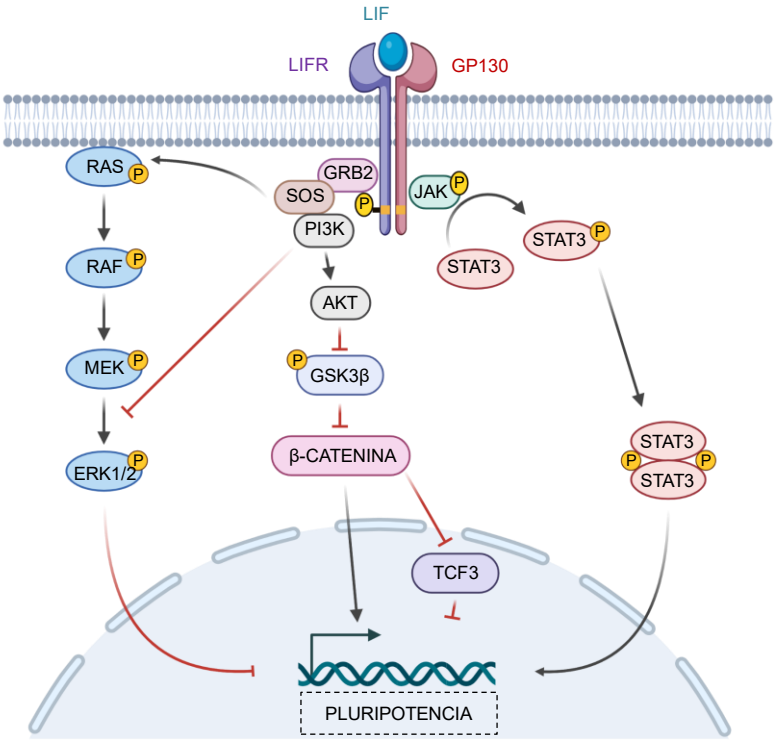


Figura I7. Vía de señalización de LIF. Representación esquemática simplificada de la señalización intracelular desencadenada en presencia del factor inhibidor de la leucemia (LIF) y los diferentes efectos sobre el estado pluripotente. Los estados fosforilados de las diversas proteínas se representan por la unión de un grupo fosfato (P). Imagen creada con BioRender.com.

Además de la activación de STAT3, la unión de LIF es capaz de estimular una segunda vía de señalización caracterizada por la activación mediante fosforilación de AKT mediada por PI3K[119,120] (**Figura I7**). AKT inhibe entonces la actividad de GSK3β a través de dos mecanismos independientes, tanto por interacción directa[120] como mediante el control de su localización subcelular[121], favoreciendo así

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

un estado pluripotente de pre-implantación al regular la expresión y niveles de *Tbx3, Nanog, Oct4 y c-Myc*[117,121-123]. Además, una vez activada la vía de PI3K, ésta es capaz de inhibir la activación de la vía MEK-ERK1/2[120] (**Figura I7**), ruta crítica para el mantenimiento y la salida de pluripotencia[124]. Por último, la señalización inducida por LIF puede activar ERK1/2. Una vez activada JAK, puede interaccionar entonces con el complejo GRB2-SOS activando así la cascada RAS-RAF-MEK-ERK1/2[124]. La acción inhibidora y activadora de la citoquina LIF sobre ERK1/2, determina la presencia de bajos niveles de activación, insuficientes para provocar la pérdida del estado pluripotente, pero sí suficientes para salir del estado alternativo de pluripotencia *naïve*.

Solamente la presencia de LIF no es suficiente para mantener el estado pluripotente de las ESCs y bloquear la diferenciación. Las proteínas morfogénicas óseas (BMP, del inglés *bone morphogenic proteins*) presentes en el suero que se añade al medio de cultivo estándar, también juegan un papel crucial a este respecto. La presencia conjunta de BMP4 y LIF mantiene el estado pluripotente y la autorrenovación indefinida de las ESCs en cultivo[125,126]. Las proteínas BMP pertenecen a la superfamilia de factores de crecimiento transformante beta (TGFβ)[127]. BMP4 se une a los receptores de la superficie celular serina/treonina quinasa de tipo I (ALK3 y ALK6) y posteriormente a los receptores de tipo II (BMPR2)[127] (**Figura I8**). Una vez activados los receptores, se fosforila y activa a las proteínas SMAD activadas por receptores (R-SMADs) SMAD1 y SMAD5, pudiendo entonces asociarse con el mediador SMAD4[128]. El complejo formado se transloca al núcleo, donde actúa como un factor de transcripción bloqueando la diferenciación de las células pluripotentes y manteniendo el estado pluripotente de pre-implantación[128,129]. Además, BMP4 es capaz de inhibir la activación de ERK1/2 mediante la activación de la fosfatasa dual específica 9 (DUSP9) por SMAD1/5-SMAD4[130] (**Figura I8**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

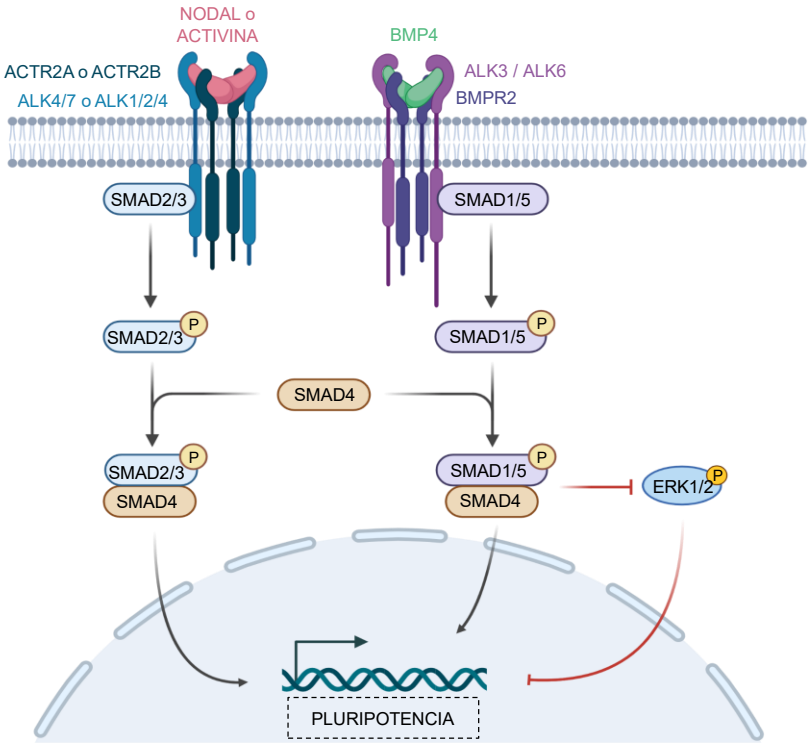


Figura I8. Vía de señalización de NODAL/ACTIVINA/BMP4. Representación esquemática simplificada de la señalización intracelular desencadenada en presencia de NODAL, ACTIVINA o BMP4 y los diferentes efectos sobre el estado pluripotente. Los estados fosforilados de las diversas proteínas se representan por la unión de un grupo fosfato (P). Imagen creada con BioRender.com.

Otros miembros de la superfamilia TGFβ son la proteína NODAL y la proteína ACTIVINA A. Al contrario que BMP, NODAL y ACTIVINA se unen a receptores tipo II (ACTR2A y ACTR2B) para luego unirse a receptores tipo I (ALK4/7 y ALK1/2/4 respectivamente)[127]. En este caso, la activación de los receptores conlleva la fosforilación y activación de SMAD2 y SMAD3, asociándose también a SMAD4, y translocándose el complejo al núcleo[131] (**Figura I8**). Al contrario que la vía mediada por BMP4, la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

finalidad de la vía NODAL/ACTIVINA es mantener el estado pluripotente post-implantación.

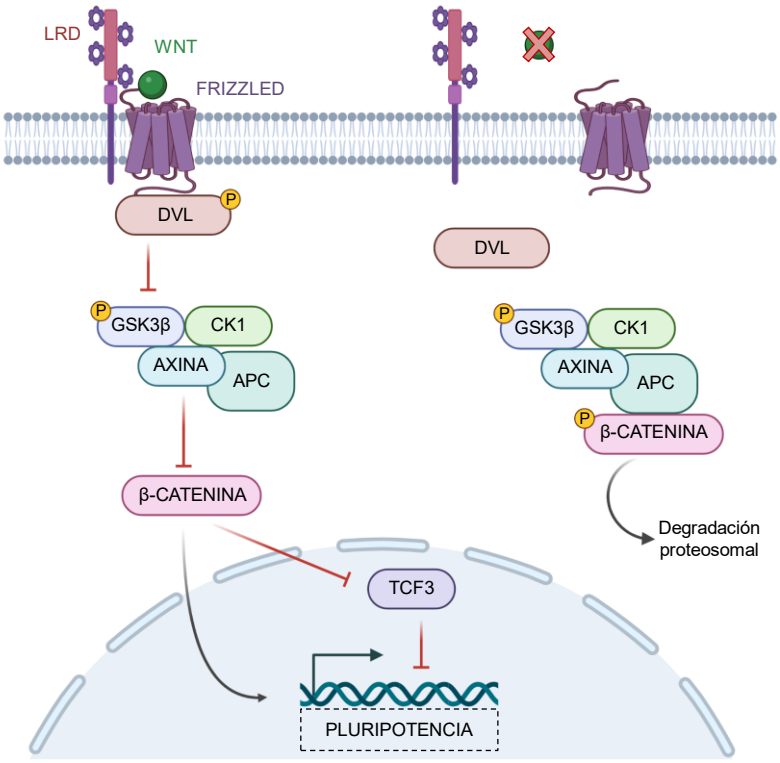


Figura I9. Vía de señalización de WNT. Representación esquemática simplificada de la señalización intracelular desencadenada en presencia o ausencia de WNT y los diferentes efectos sobre el estado pluripotente. Los estados fosforilados de las diversas proteínas se representan por la unión de un grupo fosfato (P). Imagen creada con BioRender.com.

La vía de señalización de WNT resulta esencial en diversos procesos fisiológicos, entre los que se encuentran el desarrollo embrionario y la diferenciación[132]. En ausencia de ligandos WNT, el complejo formado por la proteína de armazón AXINA, la proteína APC, la proteína quinasa CK1 y GSK3β (denominado complejo destructor) fosforila la β-CATENINA marcándola para su degradación

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

a través del proteosoma[30,31] (**Figura I9**). Cuando las proteínas WNT se unen a su complejo receptor heterodimérico FRIZZLED/LRD, permiten la unión de la proteína fosforilada DVL[31], capaz de secuestrar la AXINA, bloqueando así la acción del complejo destructor, y provocando una acumulación de β-CATENINA liberada[31]. Esta β-CATENINA liberada se transloca al núcleo, donde promueve el mantenimiento del estado pluripotente reprimiendo al factor TCF3 o de manera independiente, al favorecer la expresión y/o actividad de *Oct4*, *Sox2*, *Nanog* y *Esrrb*[133-138] (**Figura I9**).

2.2.3.2. Vía de señalización FGF-ERK1/2

Como se mencionó con anterioridad, la activación y función de ERK1/2 es crítica no solo para el mantenimiento y control de los estados pluripotentes sino también para la salida de éste y el comienzo de los procesos de diferenciación. Esta vía de señalización se modula e incrementa a medida que las células van perdiendo pluripotencia y empieza la restricción de linaje[139,140].

Los factores de crecimiento fibroblástico (FGF) son una familia de proteínas multifuncionales, con una gran variedad de efectos durante el desarrollo embrionario. Por ejemplo, se ha visto que la actividad FGF-ERK1/2 presenta un importante papel en la transición de pluripotencia *naïve* a *primed*[141,142]. Además, como se ha descrito anteriormente en esta introducción, la inhibición de ERK1/2 permite mantener en cultivo un estado celular pluripotente *naïve*[17,29], mientras que su activación es necesaria para un estado celular pluripotente *primed*[38,39]. Curiosamente, la señalización de otras vías importantes en el estado celular pluripotente puede contrarrestar la actividad de esta quinasa tal y como se ha detallado en la sección anterior. La señalización de los ligandos FGF comienza tras la unión a FGFR, que forma dímeros y se activa por autofosforilación[143]. La activación de los receptores permite la unión de FRS2, GRB2 y SOS, complejo capaz de activar RAS en la membrana citoplasmática[144]. RAS puede entonces disociarse de la membrana y activar a RAF, que a su vez activa a MEK,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

activando éste finalmente a ERK1/2[32]. La quinasa ERK1/2 activada se transloca al núcleo donde actúa como factor de transcripción, regulando procesos como la diferenciación celular[140] (**Figura I10**).

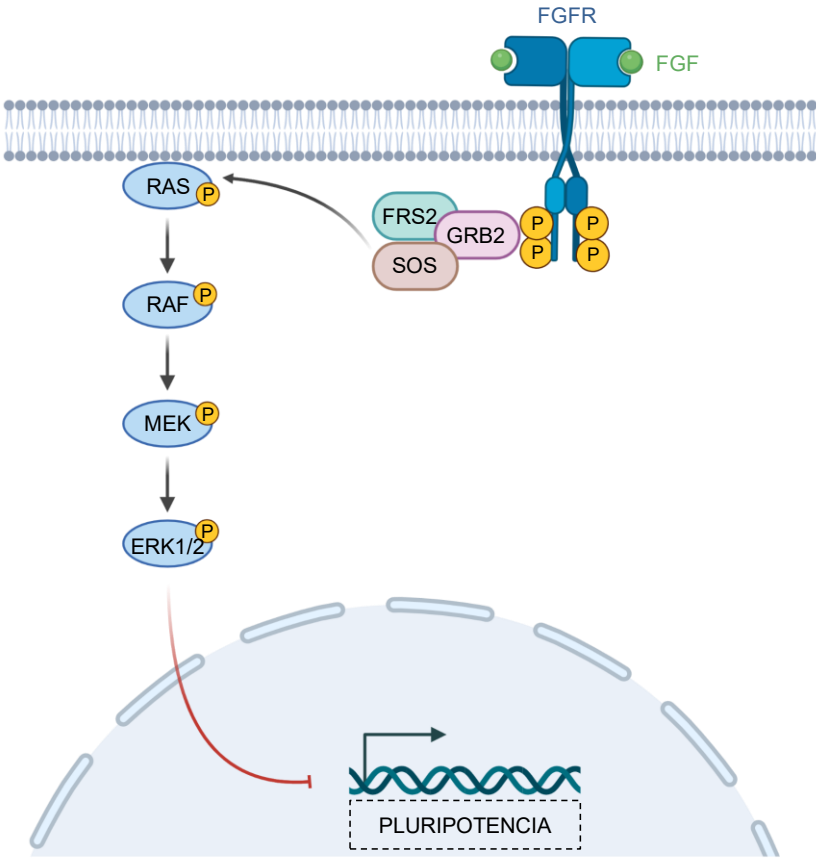


Figura I10. Vía de señalización de FGF-ERK1/2. Representación esquemática simplificada de la señalización intracelular desencadenada en presencia de FGF y su efecto sobre el estado pluripotente. Los estados fosforilados de las diversas proteínas se representan por la unión de un grupo fosfato (P). Imagen creada con BioRender.com.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.3 SALIDA DEL ESTADO CELULAR PLURIPOTENTE

Además de la capacidad de autorrenovación, las células pluripotentes poseen la capacidad de diferenciarse y dar lugar a cualquier tipo celular perteneciente al embrión. El balance entre la autorrenovación y la diferenciación es crítico para el desarrollo embrionario y la homeostasis de los tejidos. Se ha mostrado que varias quinasas controlan la red de señalización intracelular en la salida de la pluripotencia hacia distintos linajes[145]. Entre éstas, la señalización mediada por ERK1/2 ejerce un amplio impacto sobre la diferenciación y el desarrollo celular. Como hemos visto en secciones anteriores y como veremos en la sección de resultados de mi tesis doctoral, la correcta regulación de la activación de ERK1/2 es necesaria no solo para mantener la identidad de los estados alternativos pluripotentes, sino para permitir la salida y el comienzo de la diferenciación celular. Esta activación produce un rápido cambio en los niveles de los factores clave de pluripotencia NANOG, KLF4 y KLF2, reduciendo la expresión de sus genes mediante cambios en la asociación de la ARN polimerasa II a sus regiones potenciadoras[43,146]. Además, ERK1/2 inhibe directamente mediante fosforilación a OCT4, NANOG, KLF4 y KLF2, provocando en particular la exportación nuclear y la degradación de estos dos últimos factores[147-150]. No solo la activación de ERK1/2 afecta a los factores pluripotentes, sino que es necesaria para iniciar la salida de este estado. Varios estudios muestran como la inhibición de ERK1/2, ya sea mediante aproximaciones químicas o genéticas, retrasa e incluso impide la correcta diferenciación celular aun en presencia de señales específicas[105,151,152]. Asimismo, y en consonancia con los efectos descritos, el incremento en la actividad de ERK1/2 es suficiente para perder el estado pluripotente, forzando el inicio de la diferenciación celular[140].

Existe una gran cantidad de estudios en los que se describen protocolos de diferenciación de las células pluripotentes hacia los tres linajes germinales. En concreto, la especificación de endodermo se fundamenta en la hiperactivación de la vía de señalización de NODAL

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

y ACTIVINA, mientras que la determinación de mesodermo se consigue con la activación de la vía de BMP4[153-156]. Por último, la diferenciación a ectodermo se favorece mediante la exposición a ácido retinoico[157,158]. Los tratamientos combinados de las células pluripotentes con diferentes citoquinas posibilitan la obtención de tipos celulares más especializados, como por ejemplo células endoteliales (ECs, del inglés *endotelial cells*) o progenitoras endoteliales (EPCs, del inglés *endotelial progenitor cells*), las cuales fueron generadas a lo largo de mi tesis doctoral. *In vivo*, las EPCs derivan de progenitores multipotentes del mesodermo[159], y su correcta especificación a células endoteliales es crítica para el establecimiento de las redes vasculares de un organismo, ejemplificado por la letalidad durante desarrollo embrionario derivada de defectos en la misma[160,161] (**Figura I11**).

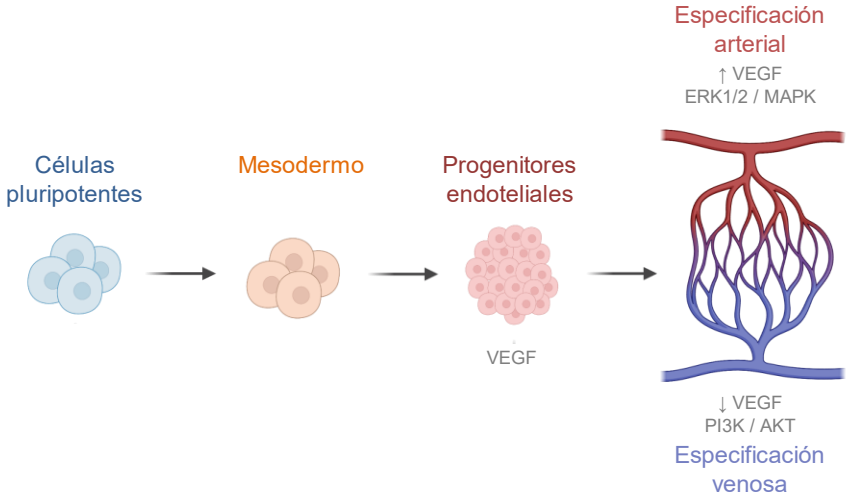


Figura I11. Diferenciación endotelial a partir de células pluripotentes. Representación esquemática simplificada de las etapas y factores más relevantes durante el desarrollo vascular. La señalización del factor de crecimiento endotelial vascular (VEGF) juega un papel crítico en la especificación endotelial, considerándose el receptor VEGFR2 marcador de progenitores endoteliales. La estimulación producida por VEGF actúa sobre las vías de ERK1/2 y PI3K para la determinación arterial o venosa[162,163]. Imagen creada con BioRender.com.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.4 PROTEÍNAS CCM

La familia de los genes *Ccm* está compuesta por los miembros *Ccm1* (también conocido como *Krit1*), *Ccm2* y *Ccm3* (también conocido como *Pdcd10*). Una característica compartida por estos genes independientes es que presentan un gran nivel de conservación entre especies[164]. Los genes *Ccm* son esenciales para el correcto funcionamiento de las células endoteliales, que como se mencionó en la sección anterior, son críticas para el establecimiento, la estructura y la función de las redes vasculares sanguíneas. De esta manera, sus defectos conllevan graves consecuencias tanto durante el desarrollo embrionario como durante la vida de un organismo adulto.

2.4.1 Estructura y función de las proteínas CCM

El gen *Ccm1* codifica para la proteína más larga (84 kDa) y mejor caracterizada de las tres. No presenta actividad catalítica y se expresa en el endotelio vascular de varios órganos, especialmente en los capilares que constituyen barreras y no en los fenestrados[165]. Dentro del tejido cerebral, CCM1 está presente en células endoteliales, en los pies de los podocitos y en las neuronas piramidales[165]. A nivel subcelular, esta proteína se localiza en el núcleo y en el citoplasma mediante la regulación de la proteína quinasa Cα[166]. CCM1 consta de un dominio amino-terminal NUDIX, seguido de tres motivos NPxY/F (aspartato-prolina-X-tirosina/fenilalanina), una región con repeticiones de anquirina (ARD, del inglés *ankyrin repeat domain*), y un dominio FERM carboxilo-terminal compuesto por tres subdominios[167,168] (**Figura I12**). A través del dominio NUDIX (que en este caso carece de la actividad hidrolasa normalmente asociada), CCM1 es capaz de unirse a la tubulina[169]. Los motivos NPxY/F permiten la interacción con ICAP1[170], CCM2[171] y SNX17[172]. Por último, el dominio FERM posibilita la unión a RAP1[173] y HEG1[174].

El gen *Ccm2* codifica una proteína de 48 kDa sin actividad catalítica, que al igual que CCM1, está presente en las células

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

endoteliales, así como en astrocitos y sus terminaciones, y neuronas piramidales del córtex[175]. A nivel subcelular, CCM2 se encuentra localizada en el citoplasma[171]. Esta proteína está constituida por un dominio de unión a fosfotirosina (PTB), un motivo similar al motivo leucina/ácido aspártico (LD-like) y un dominio de homología de harmonina helicoidal (HHD)[176] (**Figura I12**). A través de su dominio PTB interacciona con CCM1, mientras que el motivo LD-like permite la unión a CCM3, no existiendo unión directa entre CCM1 y CCM3[177]. Además, el dominio PTB también posibilita la interacción con los factores involucrados en apoptosis TRKA y SMURF1, y con la proteína de señalización RHOA[178]. Por último, mediante el dominio HHD, CCM2 puede unirse a MEKK3[179]. Se sabe que también tiene la capacidad de interaccionar con RAC1, interfiriendo con el citoesqueleto de actina, aunque aún no se ha identificado el dominio responsable[169].

El gen *Ccm3* codifica una pequeña proteína de 25 kDa sin actividad catalítica y que se expresa en el endotelio arterial y las células gliales[180]. A nivel subcelular, CCM3 se encuentra en el citoplasma y aparato de Golgi y su ausencia provoca la desestabilización de éste[181]. Esta proteína está compuesta por un dominio de dimerización (DIMER) en el extremo amino-terminal, y un dominio de homología de orientación de adhesión focal (FAT-H) en el extremo carboxilo-terminal[182] (**Figura I12**). El dominio de dimerización es el responsable de la unión de CCM3 con el complejo fosfatasa/quinasa que interactúa con la estriatina (STRIPAK) y sus componentes, representados por las STRIP1/2, la proteína fosfatasa 2A (PP2A) y la familia de quinasas serina/treonina del centro germinal (GSK) III de 20 quinasas estériles (MST4, STK24 y STK25)[183-185]. A través del domino FAT-H se une a CCM2[186], fosfotidilinositoles (PIs)[187], VEGFR2[188] y al regulador de polarización celular 1 interactor de la familia de quinasas Rho (RIPOR1)[189].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

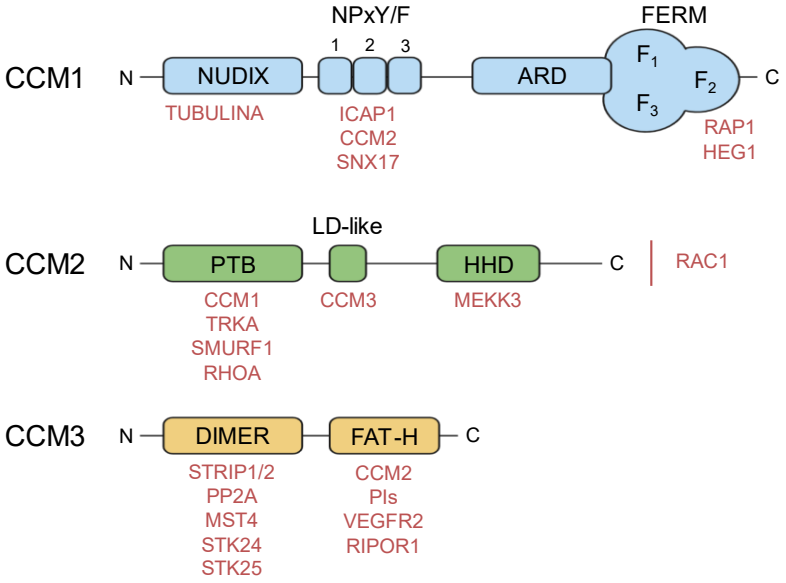


Figura I12. Dominios e interacciones de las proteínas CCM descritas en células endoteliales. Representación esquemática de la organización de CCM1 (arriba), CCM2 (centro) y CCM3 (abajo) mostrando los dominios que las conforman y las interacciones intermoleculares descritas en células endoteliales (rojo). La unión de RAC1 a CCM2 se representa a la derecha dado que el sitio concreto de unión aún no se conoce. N: extremo amino-terminal, C: extremo carboxilo-terminal. Imagen creada con BioRender.com.

Como se mencionó anteriormente, la función de las proteínas CCM es crítica para el correcto funcionamiento de las células endoteliales. CCM1, CCM2 y CCM3 pueden unirse para formar un complejo molecular de señalización, actuando CCM2 como el eje central al cual se unen tanto CCM1 como CCM3[171,190,191]. Además, las proteínas CCM pueden actuar individualmente en diferentes localizaciones subcelulares y uniéndose a diferentes proteínas (**Figura I13**). Ninguna de las tres proteínas CCM tiene dominios enzimáticos, actuando así únicamente como adaptadores[168], por lo que su efecto sobre la señalización celular opera a través de las interacciones con otras proteínas. En conjunto o individualmente, las proteínas CCM regulan la angiogénesis, el citoesqueleto de actina, las uniones célula-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

célula, la polaridad celular y la formación del lumen de los vasos sanguíneos.

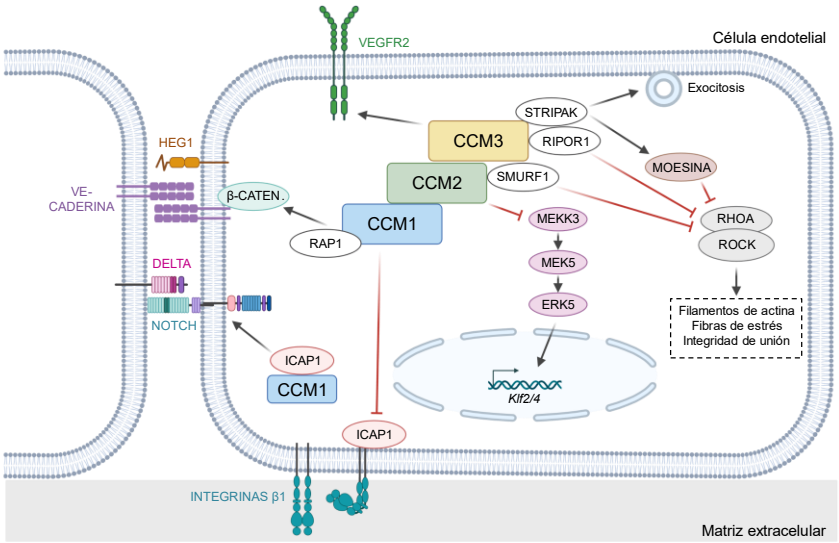


Figura I13. Funciones de las proteínas CCM en células endoteliales. Representación esquemática simplificada de las vías y procesos en los que participan CCM1, CCM2 y CCM3 directamente o a través de sus interactores. Imagen creada con BioRender.com.

Por un lado, las proteínas CCM controlan el mantenimiento de la barrera endotelial al inhibir a las quinasas RHOA y RHO[178,192], críticas para la formación de fibras de estrés, la polarización celular, la migración y la permeabilidad (**Figura I13**). CCM1 y CCM2 presentan un efecto directo sobre la actividad de las quinasas[166,193], mientras que CCM3 pertenece al complejo STRIPAK, interaccionando con la fosfatasa PP2A y las quinasas MST4, STK24 y STK25[181,183,194], que a su vez regulan al inhibidor de RHOA denominado MOESINA[195]. Asimismo, al pertenecer al complejo STRIPAK, CCM3 juega un papel principal en la regulación de la exocitosis[196,197]. Las proteínas CCM también son esenciales para la formación y mantenimiento de las

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

uniones estrechas entre células. CCM1 se une directamente a HEG1[191], encargándose de la regulación de las uniones celulares entre las células endoteliales[198,199] (**Figura I13**). Como se ha mencionado anteriormente, RAP1 interacciona con CCM1, dirigiéndola así a las zonas de contacto entre células. CCM1 es capaz entonces de unirse a HEG1, permitiendo la unión de β-CATENINA a las caderinas del endotelio vascular (VE-CADERINAS)[200]. Las uniones adherentes, estructuras basadas en caderinas, regulan diferentes procesos celulares, y juegan un papel crítico en las barreras endoteliales[201,202]. CCM2, por su parte, regula el número y localización de los componentes de las uniones y adhesiones estrechas[203] y promueve la degradación de RHOA a través de la unión con SMURF1[193] (**Figura I13**).

Por otro lado, las proteínas CCM también son críticas en los procesos de angiogénesis. La angiogénesis se define como la formación de nuevos vasos sanguíneos a partir de los ya existentes. Este proceso comprende diferentes fases como el crecimiento, la migración y la diferenciación de las células endoteliales, requiriendo una regulación exhaustiva y coordinada para que la formación de nuevos vasos se produzca cuando y donde es necesario. La pérdida de CCM1 o su interacción con ICAP1 (ITG1BP1) provoca una alteración de la vía DELTA-NOTCH, estimulando así las vías angiogénicas[204,205] (**Figura I13**). Además, ICAP1 inhibe la interacción de las integrinas β1 con la matriz extracelular al unirse a sus colas citoplasmáticas[206,207] (**Figura I13**). Las integrinas β1 también median el depósito de las fibrillas de fibronectina en la superficie celular, siendo uno de los componentes mayoritarios durante el proceso de angiogénesis por brotes[208]. CCM1 estabiliza ICAP1, controlando así la activación de las integrinas β1 y la remodelación de la fibronectina[209]. A través de CCM2, el complejo es capaz de unirse a MEKK3[171,179,210,211]. La cascada de MEKK3 interviene en la migración celular y la proliferación, y es necesaria para un desarrollo correcto de los vasos[212]. CCM2 regula negativamente a MEKK3, y en su ausencia, los niveles de la quinasa se incrementan, dando lugar a un aumento de ERK5 y finalmente de los factores KLF2

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

y KLF4[179,210,213,214] (**Figura I13**). Estos factores son críticos durante la angiogénesis y la transición de endotelio a mesénquima a través de TSP1, BMP6 y ADAMTS4[210,214]. Finalmente, CCM3 interviene en la estabilización de VEGFR2 para mantener una correcta angiogénesis[188].

2.4.2 Papel de las CCM en desarrollo embrionario

Durante el desarrollo embrionario, CCM1 se expresa en varios tejidos, entre ellos las capas celulares de neuronas, el corazón y los vasos largos, aunque no se encuentra en los vasos parenquimáticos cerebrales o de las meninges[215]. CCM2 se expresa durante la etapa de desarrollo, al igual que CCM1, en las capas celulares de neuronas, el corazón y vasos largos, aunque en este caso CCM2 sí se expresa en los vasos cerebrales y las meninges[215]. CCM3 se expresa en el hígado, riñón, pulmón y bazo del feto[216], y de manera similar a CCM1 y CCM2, se detecta también en capas celulares de neuronas, el corazón, vasos largos, vasos cerebrales y meninges[215].

La ausencia de cualquiera de las tres proteínas CCM resulta letal durante el desarrollo embrionario, lo que ha sido atribuido a fallos en los procesos angiogénicos y en el desarrollo del sistema vascular[178,188,217-219] (**Figura I14**). En ausencia de CCM1, se aprecia una dilatación arterial y estrechamiento de la arteria del primer arco branquial a día E8.5 del desarrollo, lo que limita la circulación sanguínea ya a día E9.5[217]. Asimismo, se pierde la identidad arterial, con un arresto del desarrollo a partir del día E9.5 y una retención de una red vascular primitiva en el saco vitelino[217]. Estos embriones son incapaces de completar el proceso de inversión y sufren derrames pericárdicos, dando como resultado su desintegración a día E10, incapaces de sobrevivir más allá del día E11[217]. CCM2 es crítico para la formación del lumen de los vasos de la arteria del primer arco branquial y el desarrollo del corazón. Al igual que en el caso de CCM1, la ausencia de CCM2 provoca un estrechamiento de la arteria del primer arco branquial, y a día E8.5 el embrión no exhibe circulación sanguínea[178]. Asimismo, se pierde la integridad vascular y se observa

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

la imposibilidad de organizar la red vascular del saco vitelino a día E9, produciéndose un arresto del desarrollo embrionario, derrames pericárdicos y una completa pérdida de la viabilidad a día E9.5, observándose así una reabsorción total del embrión a día E11.5[178,219]. En el caso de CCM3, su ausencia, además de provocar problemas en la angiogénesis, causa defectos en la vía de señalización de VEGFR2 y en los procesos de vasculogénesis y hematopoyesis[188]. Los embriones sin CCM3 carecen de células sanguíneas en el saco vitelino y muestran una reducción de los marcadores endoteliales[188]. Además, muestran defectos cardíacos, con ventrículos más pequeños y un miocardio más fino[188]. Al contrario que ante la ausencia de CCM1 y CCM2, el arresto del desarrollo en este caso ocurre después de la gastrulación, pero antes del inicio de la circulación, perdiéndose la viabilidad del embrión a día E8.5[188].

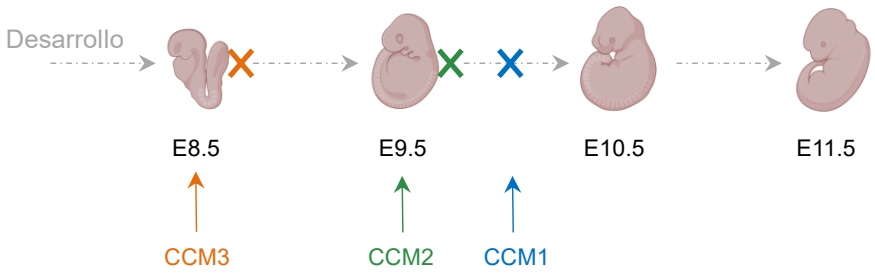


Figura I14. Relevancia de las proteínas CCM durante el desarrollo embrionario murino. Representación esquemática simplificada del intervalo del desarrollo embrionario comprendido entre los días E8.5 y E11.5 mostrando el momento en el que la pérdida de CCM1, CCM2 o CCM3 provoca la pérdida de viabilidad embrionaria. Imagen creada con BioRender.com.

2.4.3 CCMs en enfermedad

Las mutaciones de pérdida de función de cualquiera de los genes *Ccm* (*Ccm1*[220,221], *Ccm2*[222,223] o *Ccm3*[224]) causan malformaciones cavernosas cerebrales (CCM, del inglés *cerebral cavernous malformations*), una enfermedad que afecta al 0,5% de la población[225], y constituye la segunda causa más común de malformaciones

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

vasculares[226]. Esta enfermedad se caracteriza por una dilatación de los capilares, un descenso en el contacto de las células endoteliales, pérdida de proteínas de los complejos de unión y una barrera hematoencefálica comprometida, dando lugar a cavernas vasculares[227]. Se desarrolla casi exclusivamente en el sistema nerviosos central, provocando dolor de cabeza, ataques epilépticos, hemorragias, parálisis, déficits neurológicos e incluso la muerte[228,229]. Actualmente no dispone de una cura, y todo tratamiento existente es meramente paliativo. Las malformaciones se clasifican en esporádicas o familiares. Mientras que la forma esporádica aparece como lesiones solitarias, la familiar constituye una enfermedad rara (OMIM #116860) y se presenta con multitud de lesiones siguiendo un patrón de penetrancia autosómico dominante[230-233]. Para que las lesiones cavernosas se inicien, se requiere una pérdida de función de los dos alelos de uno de los genes *Ccm* en las células endoteliales maduras[234,235] o progenitoras endoteliales[236,237], tanto en la forma familiar[238,239] como en la esporádica[240] (**Figura I15**).

Este modelo, donde se requieren dos mutaciones independientes con pérdida de función de uno de los genes *Ccm* para que se manifieste la enfermedad, explica por qué las lesiones cavernosas aparecen de manera focal y por qué los individuos que padecen la enfermedad familiar presentan más lesiones que aquellos con la variante esporádica. Una vez perdida la función de los dos alelos de cualquiera de los tres genes *Ccm*, estas células progenitoras endoteliales y endoteliales maduras expresan marcadores mesenquimales y sufren una expansión clonal que termina generando los cavernomas[236,241]. A medida que estas estructuras crecen son capaces de reclutar células sanas, aumentando así la lesión y produciendo el mosaicismo característico observado en los cavernomas[235,236] (**Figura I15**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

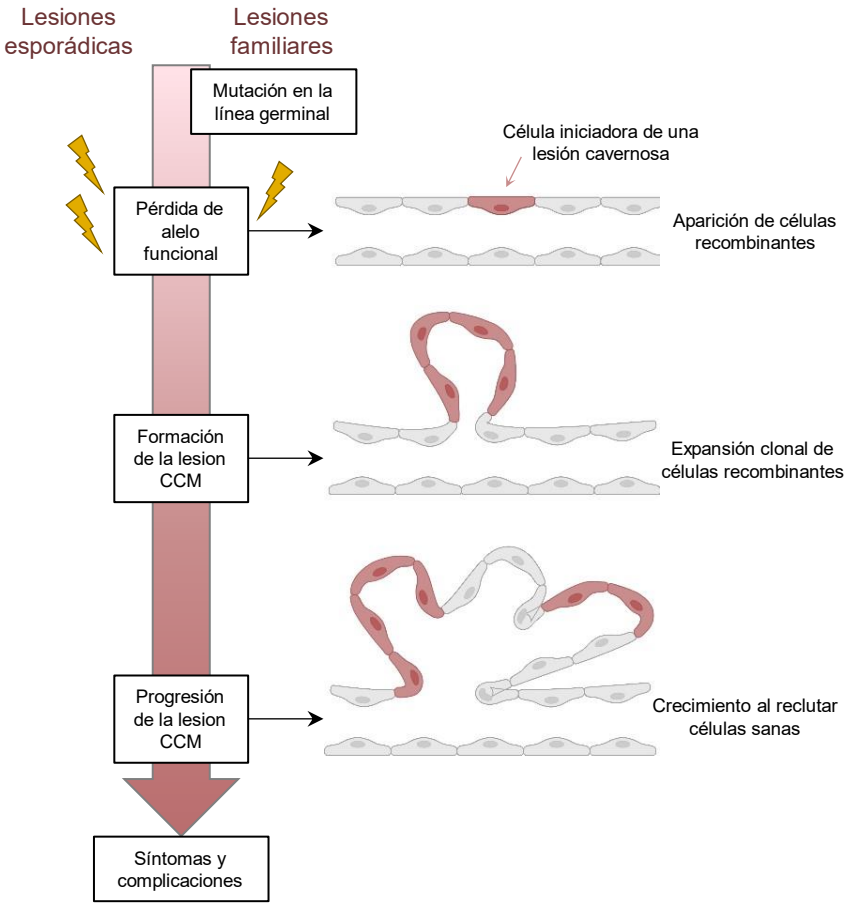


Figura I15. Modelo de la aparición y crecimiento de malformaciones cavernosas cerebrales. La pérdida de función en los dos alelos que codifican CCM1, CCM2 o CCM3 provoca la aparición de células recombinantes iniciadoras de una lesión. La expansión clonal de estas células recombinantes propicia la generación de lesiones cavernosas en el endotelio, que progresan aumentando de tamaño al reclutar células sanas cercanas. A partir de este momento pueden aparecer síntomas y complicaciones en los pacientes. Son necesarias dos mutaciones somáticas para la aparición de lesiones en casos esporádicos y solo una, además de la mutación de la línea germinal, en casos familiares. Las mutaciones somáticas se representan mediante rayos. Imagen creada con BioRender.com.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Como se ha visto en secciones anteriores, las proteínas CCM son críticas durante el desarrollo embrionario murino, resultando su ausencia letal entre los días E8.5 y E10. Puesto que no se observan diferencias morfológicas previas entre embriones con presencia o ausencia de las CCM, dicha letalidad se asocia a un funcionamiento aberrante de las células endoteliales, que impide el desarrollo y mantenimiento de una red vascular sana. Sin embargo, se desconoce su impacto durante las etapas iniciales del desarrollo embrionario y en la toma de decisiones celulares previas que determinan un proceso de gastrulación apropiado y una correcta especificación de los diversos linajes, entre los que se encuentra el endotelial.

Empleando células troncales embrionarias murinas como modelo de estudio, durante esta tesis buscaré definir el papel de cada CCM en el contexto de pluripotencia, tanto en la regulación del estado pluripotente como durante la toma de decisiones celulares que definirán las trayectorias a seguir en los primeros pasos de los procesos de diferenciación y especificación de los distintos tipos celulares. Como ejemplo de especificación, induciré la diferenciación hacia células progenitoras endoteliales y analizaré la relevancia de las proteínas CCM durante la trayectoria que determina la generación de éstas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3. OBJETIVOS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

El control del estado celular pluripotente reside en complejas redes de regulación transcripcional conservadas en las que intervienen numerosos factores de transcripción, reguladores epigenéticos y vías de señalización. Mientras que la pérdida de función de un gen *Ccm* (*Ccm1*, *Ccm2* o *Ccm3*) altera la biología endotelial, siendo suficiente para desencadenar el proceso patológico de malformaciones cavernosas cerebrales y resultando en letalidad embrionaria tras el proceso de gastrulación en ratón, la contribución de la función CCM en pluripotencia y especificación endotelial aún no se ha explorado. Con esto en mente, esta tesis doctoral se centrará en testar la hipótesis de que *la función CCM constituye una capa fundamental en la regulación de los mecanismos moleculares que gobiernan pluripotencia y que son necesarios para una correcta especificación endotelial*. Con este fin, abarcaré los objetivos principales enumerados a continuación:

Objetivo 1: identificar funciones CCM en el mantenimiento y salida del estado pluripotente.

Objetivo 2: diseccionar los mecanismos de acción específicos de cada proteína CCM.

Objetivo 3: dilucidar el impacto de las CCM en la especificación de células progenitoras endoteliales.

Espero que los hallazgos conseguidos durante mi investigación puedan contribuir a incrementar el conocimiento de las funciones regulatorias asociadas a las proteínas CCM y de los mecanismos moleculares subyacentes a la iniciación de la especificación endotelial. Además, los mecanismos identificados en este trabajo pueden resultar importantes en los contextos de desarrollo y enfermedad (p. ej. malformaciones cavernosas cerebrales) debido a su susceptibilidad de modificación a través de intervención farmacológica.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4. MATERIAL Y MÉTODOS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.1 Reactivos químicos y Kits

Todos los reactivos químicos empleados en esta tesis doctoral fueron sustancias adecuadas para su uso en aplicaciones de biología molecular. Los reactivos utilizados y sus referencias se encuentran listados en la **Tabla Suplementaria 1**. Las concentraciones utilizadas para cada aplicación específica se indican en las correspondientes secciones de material y métodos, así como en el desarrollo de los resultados. Los kits empleados a lo largo de este estudio se encuentran listados en la **Tabla Suplementaria 2**. Los accesorios y material fungible necesario para llevar a cabo todos los procedimientos se encuentran listados en la **Tabla Suplementaria 3**.

4.2 Modelos experimentales

Las líneas celulares empleadas en esta tesis se describen en la siguiente tabla, y se detallan a continuación:

Modelos Experimentales: Líneas Celulares

Nombre	Fuente	Identificación
J1 (ESCs murinas)	ATCC	SCRC-1010
J1 *Ccm1* KO (ESCs murinas)	Este estudio	n/a
J1 *Ccm2* KO (ESCs murinas))	Este estudio	n/a
J1 *Ccm3* KO (ESCs murinas)	Este estudio	n/a
Nanog-GFP (ESCs murinas)	Dr. Konrad Hochedlinger (Harvard Univ., USA)	n/a
tdTomato (ESCs murinas)	Este estudio	n/a
EpiSC-9 (EpiSCs murinas)	Paul J. Tesar (Case Western Univ., USA)	n/a
OEC2 (EpiSCs murinas)	Dr. Jose Silva (Cambridge Univ., UK)	n/a
bEnd.3 (endoteliales murinas)	ATCC	CRL-2299
HEK293T	ATCC	CRL-3216
Fibroblastos murinos	Este estudio	n/a

J1 ESCs: células pluripotentes derivadas de la masa celular interna del blastocisto de pre-implantación de ratones macho agouti 129S4/SvJae.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

J1 KO ESCs para *Ccm1*, *Ccm2* o *Ccm3*: *knock-outs* generados durante esta tesis mediante CRISPR-Cas9 (ver sección 4.9) en J1 ESCs.

Nanog-GFP ESCs: ESCs *knock-in* macho con fondo genético v6.5 y con reportero GFP bajo el control del promotor del gen *Nanog*.

tdTomato ESCs: J1 ESCs transgénicas que expresan de manera estable el reportero fluorescente tdTomato.

EpiSC-9 EpiSCs: células pluripotentes derivadas del epiblasto de un embrión murino hembra a día E5.5 con fondo genético 129SV.

OEC2 EpiSCs: células pluripotentes *knock-in* derivadas del epiblasto de un embrión murino hembra a día E5.75 con fondo genético 129SV, portadora de un reportero GFP bajo el control del potenciador del gen *Oct4*.

bEnd.3: células endoteliales derivadas del córtex cerebral de ratones BALB/c, transformada mediante infección con el vector de retrovirus NTKmT que expresa el antígeno T medio del poliomavirus.

HEK293T: células de riñón embrionario humano 293T inmortalizadas mediante la transfección del antígeno T SV40, capaz de facilitar la propagación de vectores que contienen la región replicativa SV40. Esta línea celular fue empleada para la generación de partículas lentivirales que contienen shARNs clonados en el vector pLKO (ver secciones 4.4.2 y 4.7).

Fibroblastos: cultivo primario mesenquimal derivado de oreja de ratones macho de fondo genético C57BL/6J y 4 meses de edad.

4.3 CULTIVO CELULAR

Todas los tipos celulares y el material empleado en su cultivo se mantuvieron en todo momento en condiciones de esterilidad y fue manipulado en una cabina de flujo laminar.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Para el manejo y el mantenimiento de los cultivos de todas las líneas celulares, las células fueron individualizadas con tripsina-EDTA o accutasa dependiendo de si las células son cultivadas en presencia o ausencia de suero fetal bovino (FBS), respectivamente.

Todos los cultivos celulares se realizaron a 37ºC en presencia de 5% de CO_2. Para la conservación a largo plazo, las células fueron recuperadas por centrifugación a 300 xg durante 5 minutos a temperatura ambiente (RT, del inglés *room temperature*) con el fin de eliminar cualquier resto de medio de cultivo. Tras eliminar el sobrenadante, las células se resuspendieron en medio para congelar [90% de medio de cultivo y 10% de dimetilsulfóxido (DMSO)], se dispensaron en crio-viales y se mantuvieron a -80ºC un máximo de 48 horas. Finalmente, se trasladaron a tanques de nitrógeno líquido hasta su uso.

Células troncales embrionarias (ESCs)

Las ESCs fueron cultivadas sobre placas de cultivo celular pretratadas con gelatina. Para ello, se preparó gelatina 0,1% en agua ultrapura, y se cubrió por completo la superficie de las placas de cultivo durante un mínimo de 30 minutos a 37ºC y 5% de CO_2. Tras la incubación, la gelatina fue aspirada justo antes de la siembra celular.

Las ESCs se mantuvieron en medio pluripotente estándar (SL), compuesto por medio Eagle modificado de Dulbecco alto en glucosa (DMEM), suplementado con 15% de FBS inactivado a 56ºC durante 30 minutos, 1% de mezcla de nucleósidos en agua ultrapura (adenosina, citidina, guanosina, uridina y timidina), 200 mM de L-Glutamina, 100 mM de aminoácidos no esenciales, 100 U/mL de penicilina/estreptomicina, 0,1 mM de β-mercaptoetanol y 1.000 U/mL de factor inhibidor de la leucemia (LIF; producido y testado para su uso en ESCs en el laboratorio).

Para los experimentos en los que las ESCs se cultivaron en medio de pluripotencia *naïve* (2iL), las células se mantuvieron en 50%

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de medio Eagle modificado de Dulbecco con mezcla de nutrientes F-12 Ham (DMEM/F-12) y 50% de medio Neurobasal, suplementado con 1X de B-27, 1X de N-2, 2 mM de L-Glutamina, 110 U/mL de penicilina/estreptomicina, 9,125 ng/mL de β-mercaptoetanol, 1.000 U/mL de LIF, 3 µM del inhibidor de GSK3β (CHIR99021) y 1 µM del inhibidor de MEK (PD0325901).

Células troncales del epiblasto (EpiSCs)

Las EpiSCs fueron cultivadas sobre placas de cultivo pretratadas con fibronectina. Para ello, se preparó fibronectina a una concentración de 10 µg/mL en tampón fosfato salino (PBS), y se cubrió por completo la superficie de las placas de cultivo durante 40 minutos a RT. Tras la incubación, la fibronectina fue aspirada justo antes de la siembra celular.

Las EpiSCs se mantuvieron en medio de pluripotencia *primed* (FA), compuesto por un 50% de medio DMEM/F-12 y un 50% de medio Neurobasal, suplementado con 1X de B-27, 1X de N-2, 2 mM de L-Glutamina, 110 U/mL de penicilina/estreptomicina, 9,125 ng/mL de β-mercaptoetanol, 20 ng/mL de ACTIVINA A y 12 ng/mL de FGFb.

Células endoteliales

Las células endoteliales (bEnd.3) fueron cultivadas directamente sobre placas de cultivo celular, y se mantuvieron en medio endotelial, compuesto por medio DMEM alto en glucosa suplementado con 10% de FBS inactivado a 56ºC durante 30 minutos y 1% de penicilina/estreptomicina.

Células embrionarias de riñón humanas y fibroblastos murinos

Las células embrionarias de riñón humanas (HEK293T) y los cultivos primarios de fibroblastos murinos fueron cultivados directamente sobre placas de cultivo celular, y se mantuvieron en medio somático, compuesto por medio DMEM alto en glucosa suplementado con 10%

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de suero fetal bovino inactivado a 56ºC durante 30 minutos, 200 mM de L-Glutamina y 100 U/mL de penicilina/estreptomicina.

4.4 CLONACIÓN DE VECTORES PLASMÍDICOS

4.4.1 Clonación de guías de ARN para CRISPR-Cas9

Para obtener líneas KO para *Ccm1*, *Ccm2* o *Ccm3* se empleó la herramienta de edición génica CRISPR-Cas9[242]. Se diseñaron guías de ARN contra secuencias diana específicas para el exón 1 de *Ccm1*, el exón 3 de *Ccm2* (primer exón común a las dos isoformas existentes) y el exón 2 de *Ccm3* a través de la herramienta *online* CRISPOR (http://crispor.tefor.net)[243]. Se seleccionaron las guías con mayor puntuación y menor probabilidad de unirse a otras regiones inespecíficas. Se anillaron y fosforilaron las secuencias complementarias de cada guía en una reacción compuesta por 2,5 µM de secuencia directa, 2,5 µM de secuencia reversa, 1X de tampón de reacción A de la polinucleótido quinasa T4 (PNK), 1 mM de ATP pH 7 y 10 U de T4 PNK. Seguidamente, se incubaron en un termociclador 30 minutos a 37ºC, 5 minutos a 95ºC, una reducción de temperatura de 5ºC cada minuto hasta alcanzar los 25ºC y se estabilizaron a 4ºC. Las secuencias se diseñan para que tras anillar los oligonucleótidos queden extremos cohesivos de unión a las digestiones generadas por la enzima de restricción *BbsI*. Se digirió y defosforiló el vector SpCas9-2A-Puro V2.0 (Addgene, 62988), que codifica para la enzima Cas9 y permite el clonaje de una guía, empleando *BbsI* y fosfatasa alcalina termosensible FastAP siguiendo las instrucciones del fabricante. A continuación, se aisló el vector linealizado en un gel de agarosa mediante electroforesis y se purificó empleando el kit GenElute Gel Extraction siguiendo las recomendaciones del fabricante. Finalmente, se ligaron el vector linealizado purificado y los dúplex anillados de las guías de ARN empleando la ligasa de ADN T4 y se validó su identidad mediante secuenciación Sanger. Este vector contiene genes de resistencia a ampicilina y puromicina para su selección en bacterias y mamíferos,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

respectivamente. Las secuencias de las guías de ARN empleadas en esta tesis se proporcionan en la **Tabla Suplementaria 4**.

4.4.2 Clonación de shARNs

Para silenciar la expresión de *Ccm2* en esta tesis, se emplearon ARNs silenciadores (shARNs) como se describió previamente[244]. Se diseñaron oligonucleótidos contra secuencias diana de *Ccm2* y *Luciferasa* como control, depositadas en el portal web *Broad Institute GPP* (https://portals.broadinstitute.org/gpp/public/). Se anilló cada pareja de oligonucleótidos en una reacción compuesta por 2,5 µM de oligonucleótido directo, 2,5 µM de oligonucleótido reverso y 1X de tampón de reacción para digestiones enzimáticas de New England Biolabs (NEB). Seguidamente, se incubaron en un termociclador 10 minutos a 95ºC, 10 minutos a 70ºC, una reducción de temperatura de 1ºC cada 2,5 minutos hasta alcanzar los 25ºC y se estabilizaron a 4ºC. Las secuencias se diseñan para que tras anillar los oligonucleótidos queden extremos cohesivos de unión a las digestiones generadas por las enzimas de restricción *AgeI* y *EcoRI*. Se digirió el vector pLKO.1 hygro (Addgene, 24150) con *AgeI* y *EcoRI*, se aisló el vector linealizado en un gel de agarosa mediante electroforesis y se purificó empleando el kit GenElute Gel Extraction siguiendo las instrucciones del fabricante. Finalmente, se ligaron el vector linealizado purificado y los oligonucleótidos anillados empleando la ligasa de ADN T4 y se validó su identidad mediante secuenciación Sanger. Este vector lentiviral contiene genes de resistencia a ampicilina y higromicina para su selección en bacterias y mamíferos, respectivamente. Las secuencias de los shARN empleados en esta tesis se proporcionan en la **Tabla Suplementaria 4**.

4.4.3 Clonación de regiones codificantes

Para analizar la localización y las proteínas interactoras de CCM2, se empleó el sistema de transposones PiggyBac para expresar una versión de CCM2 marcada con tres epítopos Flag en su extremo amino-terminal (^{3xFL}CCM2)[245]. La secuencia codificante (CDS) de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Ccm2 se amplificó a partir de ADN complementario (ADNc) obtenido a partir de ARN (ver secciones 4.11 y 4.12) de ESCs WT utilizando una mezcla maestra de polimerasa de alta fidelidad Platinum SuperFi II PCR y cebadores específicos para *Ccm2* que además contienen los sitios de restricción para *NheI* y *PacI*. A continuación, se aisló el amplicón del tamaño adecuado en un gel de agarosa mediante electroforesis y se purificó empleando el kit GenElute Gel Extraction siguiendo las instrucciones del fabricante. El CDS purificado se insertó en el vector pJET1.2 empleando el kit de clonaje CloneJet PCR, y se validó su identidad mediante secuenciación Sanger. A continuación, se digirieron con las enzimas de restricción *NheI* y *PacI* tanto el vector pPB-CAG-3xFLAG-pA-pgk-hph (Addgene, 48753) como el constructo generado pJET1.2-*Ccm2*. El vector linealizado y el CDS de *Ccm2* se purificaron mediante electroforesis y extracción de gel empleando el kit GenElute Gel Extraction y se ligaron empleando la ligasa de ADN T4. pPB-CAG-3xFLAG-pA-pgk-hph contiene genes de resistencia a ampicilina y higromicina para su selección en bacterias y mamíferos, respectivamente. Las secuencias de los cebadores empleados en esta tesis para la amplificación de las secuencias codificantes se proporcionan en la **Tabla Suplementaria 4**.

4.5 TRANSFORMACIÓN Y OBTENCIÓN DE VECTORES PLASMÍDICOS

4.5.1 Producción de bacterias competentes

Para generar bacterias competentes se siguió un protocolo publicado previamente[246]. Se sembró una cepa de *Escherichia coli* Stbl3 en placas de agar Luria-Bertani (LB) empleando el método de sembrado por agotamiento y se incubaron durante un máximo de 16 horas a 37ºC. A continuación, se transfirieron colonias individuales a 25 mL de medio líquido LB y se incubaron otras 16 horas a 37ºC y 225 rpm (revoluciones por minuto). Al día siguiente, se inoculó 1 mL de la solución de crecimiento bacteriano en 100 mL de medio LB y se incubó a 37ºC y 225 rpm hasta que la densidad óptica del cultivo alcanzó un valor de 0,55 de absorbancia a 600 nm. Seguidamente, el cultivo

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

bacteriano se incubó en hielo durante 10 minutos y se recogieron las bacterias mediante centrifugación a 2.500 xg durante 5 minutos a 4ºC. Las bacterias sedimentadas se resuspendieron en tampón de transformación Inoue frío compuesto por: 55 mM de $MnCl_2 \cdot 4H2O$, 15 mM de $CaCl_2 \cdot 2H2O$, 250 mM de KCl y 10 mM de PIPES pH 6.7 en agua ultrapura. Se recogieron las bacterias mediante centrifugación a 2.500 xg durante 5 minutos a 4ºC. A continuación, las bacterias sedimentadas se resuspendieron en tampón de transformación Inoue suplementado con DMSO a una concentración final del 7%, y se incubaron en hielo durante 10 minutos. Finalmente, la solución de bacterias competentes se distribuyó en alícuotas de 100 µL sometiéndolas a un congelamiento rápido empleando un baño de hielo seco-etanol. Para conservarlas a largo plazo, se almacenaron a -80ºC.

4.5.2 Transformación de bacterias competentes

Para transformar las bacterias competentes con los vectores plasmídicos generados en la sección 4.4 y poder aumentar su cantidad se siguió el protocolo descrito a continuación. Primero, se descongelaron alícuotas de bacterias competentes en hielo. Segundo, se añadió un máximo de 10% v/v de ADN plasmídico y se incubó la mezcla durante 25 minutos en hielo. Tercero, se realizó un golpe de calor incubando la mezcla durante 45 segundos a 42ºC seguido por una incubación de 5 minutos en hielo para incrementar la eficiencia del proceso de transformación. Cuarto, se añadieron 300 µL de medio SOC (compuesto por 20 g de triptona, 5 g de extracto de levadura y 0,5 g de cloruro sódico para un total de 1L), y se incubó la mezcla durante 1 hora a 37ºC y 225 rpm. Quinto, se sembraron 100-200 µL del cultivo bacteriano en placas de agar LB que contienen antibiótico ampicilina y se incubaron durante 16 horas a 37ºC.

4.5.3 Obtención del ADN plasmídico

Para extraer y purificar los vectores plasmídicos amplificados mediante la transformación de bacterias competentes se siguió el protocolo descrito a continuación. Se cultivaron colonias individuales

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

crecidas en placas de agar LB en 8 mL o 100 mL de medio líquido LB suplementado con 100 µg/mL de antibiótico ampicilina durante 16 horas a 37ºC y 225 rpm, según se necesitara mayor o menor cantidad de ADN. A continuación, se extrajo y purificó el ADN plasmídico mediante el kit GenElute Plasmid Miniprep (cultivos de 8 mL) o PureLink HiPure Plasmid Filter Midiprep (cultivos de 100 mL) siguiendo las instrucciones de los fabricantes. La calidad y cantidad de ADN se determinó empleando un espectrofotómetro Nanodrop 2000.

4.6 TRANSFECCIÓN DE CÉLULAS DE MAMÍFERO

Para la expresión de ^{3xFL}CCM2 basado en el sistema PiggyBac se transfectaron ESCs usando el kit jetPRIME con el vector generado en la sección 4.4.3 y un vector que contiene la transposasa (pBASE) para activar el sistema (proporcionado por José C. Silva, Cambridge University, UK). Se sembraron 115.000 ESCs WT por pocillo de una placa de 12 pocillos en medio pluripotente estándar de cultivo (SL). Al día siguiente, se añadió medio fresco y se transfectaron las células añadiendo gota a gota una mezcla de 0,4 µg de los constructos generados y 0,4 µg de pBASE en 75 µL tampón jetPRIME conteniendo 1 µL de reactivo jetPRIME, incubada previamente durante 10 minutos a RT. Cuatro horas después se cambió el medio por SL fresco para aumentar la eficiencia de transfección. Veinticuatro horas después se comenzó la selección con 250 µg/mL de antibiótico higromicina. Se mantuvo la selección añadiendo medio fresco cada día hasta que un control con ESCs sin transfectar presentó 100% de muerte celular.

Para la expresión del constructo generado para la obtención de líneas KO mediante CRISPR-Cas9 (ver sección 4.4.1) se transfectaron ESCs WT de manera similar empleando el kit jetPRIME. En este caso, se utilizaron 0,8 µg del ADN plasmídico correspondiente y se seleccionaron las células transfectadas durante dos días con 2 µg/mL de antibiótico puromicina.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.7 GENERACIÓN DE PARTÍCULAS LENTIVIRALES

Los constructos lentivirales pLKO generados en la sección 4.4.2 se encapsularon en partículas virales empleando protocolos publicados previamente[247,248]. Para la transfección, se sembraron $8x10^6$ HEK293T en placas de cultivo de 150 mm de diámetro en 30 mL de medio somático (ver sección 4.3). Tras 24 horas, se transfectaron las células empleando polietilenimina (PEI) y las siguientes soluciones:

- Solución con PEI: se añadieron 16 µL de PEI 5X a un total de 1,6 mL de solución de NaCl 150 mM. Esta solución se mantuvo en hielo hasta su uso.
- Solución con ADN: se añadieron 10 µg de cada uno de los vectores empaquetadores psPAX2 (Addgene, 12260) y pMDG2 (Addgene, 12259) y 20 µg de constructo lentiviral pLKO generado a un total de 1,6 mL de solución de NaCl 150 mM.

A continuación, la solución con PEI se añadió gota a gota sobre la solución con ADN, se mezcló mediante agitación con vórtex y se incubó durante 30 minutos a RT. Durante esta incubación, se cambió el medio de las HEK293T por medio fresco. Finalizada la incubación, se añadió gota a gota la mezcla de PEI-ADN a las células. Al día siguiente se cambió el medio de cultivo, y las partículas lentivirales se recogieron y concentraron a las 48 y 72 horas tras la transfección. Para esto, se recogió el medio de cultivo, se centrifugó durante 5 minutos a 300 xg para sedimentar cualquier resto celular y se trasvasó el sobrenadante a filtros Amicon Ultra-15 para concentrar las partículas lentivirales mediante centrifugación durante 30 minutos a 3.500 xg a 4ºC en un rotor basculante. Los sobrenadantes concentrados se almacenaron a -80ºC en alícuotas para evitar repetidos ciclos de congelación-descongelación hasta su uso.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.8 TRANSDUCCIÓN DE CÉLULAS DE MAMÍFERO

Las células pluripotentes se transdujeron mediante partículas lentivirales siguiendo un protocolo publicado previamente con pequeñas modificaciones[248]. Brevemente, se sembraron 15.000 células/cm^2 de ESCs en placas tratadas con gelatina 0,1% o EpiSCs en placas tratadas con fibronectina 10 µg/mL. Se preparó un pocillo por cada shARN y dos pocillos extra como controles positivo y negativo para la selección con antibiótico. Al día siguiente, se añadió medio fresco suplementado con 8 µg/mL de polibreno para facilitar la transducción y 50-75 µL de las correspondientes partículas virales concentradas (ver sección 4.7). Tras 24 horas, se lavaron los cultivos con PBS y se tripsinizaron las células para su selección con 250 µg/mL de antibiótico higromicina, excepto el control negativo de selección. Se mantuvieron los cultivos añadiendo antibiótico y medio fresco cada día hasta obtener una mortalidad del 100% en el control positivo de selección.

4.9 GENERACIÓN DE LÍNEAS *KNOCK-OUT* MEDIANTE CRISPR-CAS9

Con el fin de obtener líneas de ESCs *knock-out* (KO) para *Ccm1*, *Ccm2* o *Ccm3* se empleó la herramienta de edición génica CRISPR-Cas9 como se describió anteriormente[242]. Se transfectaron células troncales embrionarias J1 con el vector que contiene el ARN guía, la Cas9 y el gen de resistencia a puromicina generado en la sección 4.4.1 siguiendo el protocolo descrito en la sección 4.6. Tras dos días de selección con el antibiótico puromicina se levantaron las células y se sembraron a densidad clonal para permitir el aislamiento manual de colonias individuales procedentes de una única célula. Se extrajo ADN genómico (ver sección 4.10) de los clones aislados y se identificaron aquellos clones con genotipo KO siguiendo una estrategia de genotipado por PCR. Brevemente, se combinaron dos cebadores directos diferentes, cada uno de los cuales se une a su correspondiente alelo *wildtype* (WT) o mutante, con un cebador reverso, dando como resultado amplicones que difieren en sus tamaños finales. Se preparó

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

una mezcla de ADN genómico, la mezcla maestra de polimerasa de ADN DreamTaq Green PCR y 1 µM de cebadores directos y reversos específicos para las regiones génicas de interés. La reacción se llevó a cabo en un termociclador SimpliAmp empleando las siguientes condiciones:

Ciclos	Temperatura (ºC)	Tiempo
1	95	3 minutos
35	95	15 segundos
	Ccm1 y *Ccm2*: 58 *Ccm3: 56*	15 segundos
	72	1 minuto
1	72	10 minutos
	4	mantener

A continuación, las amplificaciones se visualizaron mediante electroforesis en un gel de agarosa empleando un transiluminador Gel Doc XR+ Molecular Imager con el software Image Lab (Bio-Rad). De cada clon KO candidato se extrajo ARN (ver sección 4.11) y proteína (ver sección 4.14.1) y se validaron aquellos KO verdaderos mediante secuenciación Sanger, análisis de la expresión génica mediante PCR cuantitativa (ver sección 4.13) y análisis de la presencia proteica mediante western blot (ver sección 4.15). La identificación de las deleciones a partir de la secuenciación Sanger se realizó mediante el software Sequencher (v5.4). La simulación de la proteína alterada se realizó mediante el software SerialCloner (v2.6.1). Las secuencias de los cebadores empleados en esta tesis se proporcionan en la **Tabla Suplementaria 4**.

4.10 EXTRACCIÓN Y PURIFICACIÓN DE ADN GENÓMICO

Para aislar ADN genómico se obtuvieron 10^4-10^6 células de las líneas ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* y se empleó el kit Centra Perugene siguiendo las instrucciones del fabricante. La calidad y cantidad de ADN se determinaron empleando un espectrofotómetro Nanodrop 2000. Las muestras se mantuvieron a -20ºC hasta su uso.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.11 EXTRACCIÓN Y PURIFICACIÓN DE ARN

Para los análisis de expresión génica mediante PCR cuantitativa y la obtención de ADN complementario se purificó ARN empleando el kit 1 E.Z.N.A Total RNA siguiendo las instrucciones del fabricante y se almacenaron las muestras a -80ºC hasta su uso. Tras la purificación se analizó la calidad y cantidad de ARN empleando un espectrofotómetro Nanodrop 2000.

Para la secuenciación masiva de ARN (RNA-seq), se sembraron $5x10^5$ células en pocillos de placas de 6 pocillos en duplicado. A las 48h, se lisó las células con el reactivo TRIzol y se procedió a purificar el ARN total empleando dicho reactivo, así como tubos Phasemaker, siguiendo las recomendaciones de los fabricantes. La cantidad y calidad de ARN se analizó mediante un fluorómetro Qubit empleando los kits Qubit HS Assay y Qubit RNA IQ Assay, respectivamente. La integridad del ARN se visualizó mediante electroforesis corriendo 500 ng de cada ARN extraído en un gel de agarosa al 1%.

4.12 OBTENCIÓN DE ADN COMPLEMENTARIO

El ADN complementario (ADNc) empleado para análisis de expresión génica mediante PCR cuantitativa y amplificación de secuencias codificantes (CDS) se obtuvo a partir de ARN total, empleando qScript cDNA SuperMix. Esta mezcla comercial contiene cebadores hexaméricos aleatorios y cebador oligo (dT), permitiendo la retrotranscripción de todos los ARN celulares. Se almacenaron las muestras a -20ºC hasta su uso.

El ADNc empleado para análisis de expresión génica mediante PCR y electroforesis se obtuvo a partir de una cantidad determinada de ARN total empleando SuperScript IV RT. Las muestras de ARN se mezclaron con dATP, dCTP, dGTP y dTTP a una concentración final de 0,5 mM cada uno y cebadores hexaméricos aleatorios a una concentración final de 2,5 µM. Se incubaron las muestras 5 minutos a 65ºC y se añadió tampón SSIV a una concentración final de 1X, DDT

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

a una concentración final de 5 mM, 20 U de RiboLock y 200 U de la reverso transcriptasa SuperScript IV. A continuación, se incubaron en un termociclador 10 minutos a 50ºC, 5 minutos a 85ºC y se estabilizaron a 4ºC. Se almacenaron las muestras a -20ºC hasta su uso.

4.13 ANÁLISIS DE EXPRESIÓN GÉNICA MEDIANTE QPCR O PCR

Para analizar y validar la expresión génica mediante PCR cuantitativa (qPCR) se siguió el protocolo descrito a continuación. Se preparó una mezcla con ADNc (ver sección 4.12), la mezcla maestra PowerUp SYBR Green y 1 µM de cebadores directos y reversos específicos para las regiones de los genes de interés. Las mezclas se cargaron en placas de reacción de 384 pocillos y se sellaron con películas PlateMax UltraClear. La reacción se llevó a cabo en el sistema para qPCR QuantStudio 5 empleando las siguientes condiciones:

Ciclos	Temperatura (ºC)	Tiempo
1	98	10 minutos
40	95	15 segundos
	55	15 segundos
	72	40 segundos
1	95	5 segundos
1	65 (subiendo 5ºC/minuto hasta alcanzar 97ºC)	1 minuto
1	97	15 segundos
-	4	mantener

Los datos generados se analizaron siguiendo el método delta-delta Ct a partir de triplicados, y los valores de expresión relativa se normalizaron con respecto a genes control y la correspondiente muestra WT o de referencia. Todas las figuras se generaron empleando el programa GraphPad Prism 9. Las secuencias de los cebadores empleados en esta tesis se proporcionan en la **Tabla Suplementaria 4**.

Para analizar y validar la expresión génica de *Flk1* y *U6* mediante PCR y electroforesis se preparó una mezcla con ADNc (ver sección 4.12), la mezcla maestra de polimerasa de ADN DreamTaq

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Green PCR y 1 µM de cebadores directos y reversos específicos para las regiones de ambos genes. La reacción se llevó a cabo en un termociclador SimpliAmp empleando las mismas condiciones descritas arriba. A continuación, las amplificaciones se visualizaron mediante electroforesis en un gel de agarosa al 2,5% empleando un transiluminador Gel Doc XR+ Molecular Imager con el software Image Lab (Bio-Rad). Se cuantificaron los niveles de expresión de *Flk1* relativos al control *U6* empleando el software Image J (v1.33).

4.14 EXTRACCIÓN Y PURIFICACIÓN DE PROTEÍNA

4.14.1 Proteína total

Para obtener lisados de proteína total se empleó el sistema de tampón de lisis RIPA siguiendo las instrucciones del fabricante. Brevemente, se recogieron células en cultivo, se centrifugaron 5 minutos a 300 xg y se resuspendieron en PBS para su lavado. A continuación, se recogieron otra vez mediante centrifugación y se resuspendieron en tampón de lisis RIPA con inhibidores de proteasas y fosfatasas recién añadidos. Se incubaron las muestras 30 minutos en hielo y se mezclaron empleando un agitador vórtex cada 2-3 minutos. Finalmente, se centrifugaron las muestras durante 10 minutos a 10.000 xg a 4ºC y se almacenó el sobrenadante (lisado total) a -80ºC hasta su uso.

4.14.2 Fracciones proteicas subcelulares

Para obtener fracciones proteicas citoplasmáticas y nucleares se siguió un protocolo descrito anteriormente[249]. Brevemente, se recogieron células en cultivo, se centrifugaron 5 minutos a 300 xg y se resuspendieron en PBS para su lavado. A continuación, se recogieron otra vez mediante centrifugación, se calculó el volumen de las células sedimentadas (PCV, del inglés *packed cell volumen*) y se resuspendieron en 5 volúmenes de tampón de extracción nuclear A frío (ver a continuación) con 0,5 mM de DTT, 0,2 mM de PMSF y 0,1 mM del cocktail inhibidor de proteasas (PIC) recién añadidos. Se

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

centrifugaron las muestras durante 5 minutos a 2.560 xg a 4ºC y se recogió el sobrenadante como fracción citoplasmática. A continuación, los núcleos sedimentados se suspendieron en 3 volúmenes de tampón de extracción nuclear A frío como anteriormente y se incubaron las muestras durante 10 minutos en hielo. Seguidamente se homogenizaron las muestras mediante el paso por agujas de 18G ½. Se centrifugaron durante 5 minutos a 4.300 xg a 4ºC y se resuspendieron los núcleos sedimentados en tampón de extracción nuclear C frío (ver a continuación) con inhibidores de proteasas, pero sin DTT. Se homogenizaron las muestras mediante el paso por agujas de 21G ½ y se incubaron durante 30 minutos a 4ºC bajo agitación en rotación. Finalmente, se centrifugaron las muestras durante 40 minutos a 18.000 xg a 4ºC para sedimentar el material insoluble y se recogió el sobrenadante como fracción nuclear. Tanto la fracción citoplasmática como la nuclear se almacenaron a -80ºC hasta su uso.

- Tampón de extracción nuclear A: 10 mM de HEPES pH 7,9, 1,5 mM de $MgCl_2$ y 10 mM de KCl en agua ultrapura.
- Tampón de extracción nuclear C: 20 mM de HEPES pH 7,9, 25% de glicerol, 0,42 M de NaCl, 1,5 mM de $MgCl_2$ y 0,2 mM de EDTA en agua ultrapura.

Para obtener fracciones proteicas citoplasmáticas, solubles nucleares y unidas a cromatina se empleó un kit de fraccionamiento subcelular de proteínas siguiendo las instrucciones del fabricante. Brevemente, se recogieron células en cultivo y se contaron. Se prepararon alícuotas de $9x10^6$ células, se centrifugaron 5 minutos a 300 xg y se resuspendieron en PBS para su lavado. A continuación, se recogieron otra vez mediante centrifugación y se procedió a comenzar el protocolo de extracción. Las fracciones proteicas se almacenaron a -80ºC hasta su uso.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.15 ANÁLISIS MEDIANTE WESTERN BLOT

Para analizar los niveles de proteínas específicas tanto en los lisados totales como en las distintas fracciones a través de la técnica western blot se siguió el protocolo descrito a continuación. Se cuantificó la cantidad de proteína en las distintas muestras empleando el kit de ensayo de proteínas BCA Pierce siguiendo las indicaciones del fabricante.

Las proteínas se prepararon añadiendo 1X de tampón de carga Laemmli (ver a continuación) y se desnaturalizaron mediante incubación durante 5 minutos a 95ºC. Las distintas muestras se cargaron para su resolución mediante electroforesis en geles Novex WedgeWell 4–20% tris-Glycine Gel o geles de acrilamida/bisacrilamida preparados en el laboratorio. A continuación, se transfirieron a membranas de fluoruro de polivinilideno durante toda la noche a 50 V a 4ºC. Al día siguiente las membranas se lavaron en agua ultrapura y se tiñeron con reactivo de Ponceau S durante 5 minutos en agitación suave. Seguidamente se lavaron 3 veces durante 5 minutos en agitación en TBS-T, consistente en tampón salino Tris 1X suplementado con 0,1% de TWEEN 20. Las membranas se bloquearon para evitar uniones inespecíficas en leche en polvo 5% diluida en TBS-T durante 1 hora a RT en agitación suave y se lavaron 3 veces durante 5 minutos en TBS-T como anteriormente. A continuación, las membranas se incubaron con los correspondientes anticuerpos primarios diluidos en 0,5% de leche en TBS-T en agitación suave y se lavaron 3 veces durante 5 minutos en TBS-T. Tras esto, se incubaron 45 minutos a 4ºC en agitación con los correspondientes anticuerpos secundarios conjugados con HorseRadish Peroxidase (HRP) para su detección mediante quimioluminiscencia. Finalmente, las membranas se lavaron 3 veces durante 10 minutos en TBS-T y se incubaron con el sustrato para inmunotransferencia ECL Pierce. La señal HRP se detectó a través de películas de rayos X Fuji y el procesador de películas Fuji FPM-100A. Se cuantificaron los niveles proteicos relativos al control empleando el

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

software Image J (v1.33). Los anticuerpos y sus condiciones de uso empleados en esta tesis se proporcionan en la **Tabla Suplementaria 5**.

- Tampón de carga Laemmli 4X: 8% de SDS, 40% de glicerol, 5% de β-mercaptoetanol, 0,04% de bromofenol azul y 240 mM de Tris-HCl pH 6,8 en agua ultrapura filtrado a través de un filtro de 0,22 µm empleando una jeringuilla.

4.16 TINCIÓN DE LA FOSFATASA ALCALINA

Las células pluripotentes muestran presencia de altos niveles de la enzima fosfatasa alcalina (AP, del inglés *alkaline phosphatase*) comparado con células somáticas, por lo que se usa rutinariamente como marcador de pluripotencia[250]. La presencia de AP se determinó empleando el kit Leucocyte Alkaline Phosphatase siguiendo las instrucciones del fabricante. Se lavaron células en cultivo con PBS con Ca^{+2} y Mg^{+2} y se fijaron durante 30 segundos a RT con la solución de fijación recién preparada. Seguidamente, se lavaron las células con agua desionizada durante 45 segundos a RT y se incubaron con la solución de tinción recién preparada durante 15 minutos a RT en oscuridad. Finalmente, las células se lavaron nuevamente con agua desionizada durante 2 minutos a RT y se dejaron secar al aire. Las imágenes se obtuvieron empleando un microscopio invertido Olympus IX51 con una cámara Olympus DP72.

4.17 INMUNOFLUORESCENCIA

Para analizar los niveles de expresión y localización de diversas proteínas mediante inmunocitoquímica se sembraron $1x10^4$ ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* por pocillo en una placa de 48 pocillos en medio pluripotente estándar de cultivo (SL). Al día siguiente se fijaron las células con 300 µL de paraformaldehido 4% durante 15 minutos a RT en oscuridad. Seguidamente, se lavaron 3 veces durante 5 minutos en 500 µL de PBS y se permeabilizaron y bloquearon las células en un único paso incubándolas durante 30 minutos a 37ºC en 300 µL de tampón compuesto por 5% de BSA y 0,3% de Triton X-100

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

en PBS. Se retiró el tampón y se lavaron una vez las células con PBS, seguidamente se incubaron durante toda la noche a 4ºC con los correspondientes anticuerpos primarios diluidos en tampón compuesto por 1% de BSA y 0,3% de Triton X-100 en PBS. Al día siguiente se lavaron las células 3 veces durante 5 minutos en 500 µL de PBS y se incubaron durante 1 hora a RT en oscuridad con los correspondientes anticuerpos secundarios conjugados con fluoróforos en el mismo tampón y se añadió 1 µM de 4',6-diamidino-2-phenylindole (DAPI) para la detección de los núcleos celulares. Finalmente, las células se lavaron 3 veces durante 5 minutos en 500 µL de PBS y se conservaron en PBS a 4ºC en oscuridad. Las imágenes de inmunofluorescencia se tomaron empleando el microscopio invertido Leica DMI6000 B con una magnificación 40X, con ajustes constantes para cada fluoróforo individual a través de las distintas condiciones experimentales. Las imágenes se procesaron entonces empleando el software de Leica LAS X (v3.4.2.18368). Los anticuerpos y sus condiciones de uso empleados en esta tesis se proporcionan en la **Tabla Suplementaria 5**.

4.18 ANÁLISIS DE CICLO CELULAR

Para analizar el ciclo celular de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* mediante tinción con yoduro de propidio se siguió el protocolo descrito a continuación. Se sembró el mismo número de células para todas las condiciones sobre gelatina 0,1% en medio estándar (SL). Al día siguiente se recogieron las células, se centrifugaron 5 minutos a 300 xg y se resuspendieron en PBS para su lavado. Seguidamente, se fijaron con etanol 70% frío, añadido por goteo mientras se agitaban las muestras a 16 rpm, y se incubó durante 30 minutos a 4ºC. Tras la fijación, las células se atemperaron 5 minutos a RT y se centrifugaron 5 minutos a 500 xg. Las células sedimentadas se resuspendieron 2 veces en PBS para su lavado centrifugándose 5 minutos a 500 xg. A continuación, las células sedimentadas se resuspendieron en solución de tinción [1X de yoduro de propidio y 10 µg/mL de ribonucleasa A en PBS] y se incubaron durante 30 minutos a 37ºC en oscuridad.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Finalmente, la intensidad del yoduro de propidio se analizó mediante citometría de flujo (ver sección 4.19).

4.19 CITOMETRÍA DE FLUJO

Para analizar la intensidad de fluorescencia de los reporteros GFP y tdTomato o del yoduro de propidio mediante citometría de flujo activada por fluorescencia (FACS) se recogieron las células, se centrifugaron durante 5 minutos a 300 xg y se lavaron con PBS. Tras el lavado las células se resuspendieron en PBS y se mantuvieron en hielo en tubos de citometría hasta su análisis. La intensidad de fluorescencia se midió a través de un citómetro FACSCalibur (BD Biosciences). Se analizaron un total de 10.000 células por condición estableciendo los parámetros determinados según el espectro de absorción/emisión de cada marcador fluorescente empleado. Los datos generados se analizaron mediante el software FlowJo (v7.6.1). Para el análisis con yoduro de propidio el tratamiento de las muestras se describe en la sección 4.18.

4.20 ENSAYOS DE PROLIFERACIÓN

4.20.1 Proliferación por contaje

Se analizaron los patrones de proliferación de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en condiciones estándar de cultivo (SL) como se describe a continuación. Se sembraron $5x10^4$ células/cm^2 de cada condición sobre gelatina 0,1%. Tras 2 días en cultivo, se recogieron las células, se resembró un porcentaje en nuevos pocillos para continuar el ensayo y se utilizó el resto para contar el número de células totales empleando una cámara de Neubauer. Este proceso se repitió cada 2 días durante un total de 6, calculando el número total de células acumuladas a lo largo del ensayo.

4.20.2 Ensayo de competición de la proliferación

Se analizaron los patrones de proliferación de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en condiciones de cultivo *naïve* (2iL) o

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

primed (FA) como se describe a continuación. Se preparó un co-cultivo de cada condición a analizar con ESCs que expresan constitutivamente el reportero fluorescente tdTomato en una proporción 4:1 (80% células WT o KO + 20% células tdTomato positivas). Se sembraron sobre gelatina 0,1% las células en condiciones *naïve* y sobre fibronectina 10 μg/mL las células en condiciones *primed*. Al día siguiente se recogieron las células, se resembró un porcentaje en nuevos pocillos para continuar el ensayo y se utilizó el resto para analizar el porcentaje de células tdTomato negativas y positivas mediante FACS (ver sección 4.19). Se repitió el mismo proceso tras 4 días de cultivo.

4.21 Ensayos de formación de colonias pluripotentes

Para analizar las consecuencias de la pérdida de función de CCM1, CCM2 o CCM3 sobre la capacidad de autorrenovación de las células se llevó a cabo un ensayo de formación de colonias siguiendo el protocolo descrito a continuación. Se sembraron ESCs o células diferenciadas WT o KO para *Ccm1*, *Ccm2* o *Ccm3* a baja densidad (80 células/cm^2) sobre gelatina 0,1% o fibronectina 10 μg/mL en medio pluripotente estándar de cultivo (SL). Al día siguiente, se cambió el medio por SL fresco, medio con suero sin LIF, medio *naïve* (2iL) o medio *primed* (FA). Los ensayos en medio *primed* se realizaron sobre fibronectina, el resto de las condiciones sobre gelatina. Se añadió medio fresco cada 2 días, y tras 5 días en cultivo se realizó una tinción de la fosfatasa alcalina (ver sección 4.16). Se contaron las colonias y se clasificaron según su tinción en: colonias no diferenciadas (morfología compacta y tinción homogénea), colonias parcialmente diferenciadas (morfología irregular y tinción heterogénea) y colonias diferenciadas (morfología plana y pérdida completa de tinción).

4.22 Análisis de los niveles de *Nanog*-GFP

Para determinar la homogeneidad en la expresión de NANOG de los cultivos de ESCs se empleó una línea celular que expresa la proteína fluorescente verde GFP bajo el control del gen *Nanog* (*Nanog*-GFP).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Brevemente, se transdujeron células *Nanog*-GFP mediante partículas lentivirales conteniendo shARNs contra *Ccm2* o *Luciferasa* como control siguiendo el procedimiento descrito anteriormente (ver sección 4.8). Se recogieron las células, se centrifugaron durante 5 minutos a 300 xg y se lavaron con PBS. Tras el lavado, se resuspendieron las células en PBS y se mantuvieron en hielo en tubos de citometría hasta su análisis mediante citometría de flujo siguiendo el procedimiento descrito anteriormente (ver sección 4.19).

4.23 INMUNOPRECIPITACIÓN

Para analizar la unión de proteínas de interés a CCM2 se realizó una extracción de proteína total de ESCs *Ccm2* KO +^{3xFL}CCM2 seguida de una inmunoprecipitación de FLAG empleando el protocolo descrito a continuación. Se creció una placa de 150 mm de diámetro hasta alcanzar una confluencia del 80-90%. Se recogieron las células, se centrifugaron durante 5 minutos a 300 xg y se lavaron con PBS frío. Seguidamente, se recogieron otra vez mediante centrifugación, y se lisaron las células en tampón de lisis T (ver a continuación) durante 30 minutos a 4ºC. Se sonicaron las muestras empleando el sonicador digital Sonifier de Branson aplicando 3 ciclos (1 ciclo: 30 repeticiones de 1 segundo ON y 0,5 segundos OFF) con una amplitud del 25% manteniendo siempre las muestras en un baño de agua-hielo. Seguidamente, se centrifugaron las muestras durante 20 minutos a 18.000 xg a 4ºC. Los lisados se almacenaron a -80ºC hasta su uso.

Para la inmuprecipitación (IP) se cuantificó la proteína empleando el kit de ensayo de proteínas BCA Pierce siguiendo las indicaciones del fabricante. A continuación, se añadió 1 mg de proteína y 2 µg de anticuerpo primario contra FLAG a tubos de baja adherencia Protein LoBind, ajustando la reacción a un total de 1 mL con tampón de lisis T, y se incubaron durante toda la noche a 4ºC bajo agitación en rotación. Al día siguiente, se prepararon 600 µg de bolas magnéticas (Dynabeads Protein G) por cada inmuprecipitación en tampón de lisis T y se realizaron 2 lavados con el mismo tampón empleando el imán

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

DynaMag-2. Seguidamente, se resuspendieron las bolas en 200 µL de tampón de lisis T por reacción, se añadieron a las mezclas de proteína + anticuerpo y se incubaron durante 3 horas a 4ºC bajo agitación en rotación. Posteriormente, se recogieron las mezclas de bolas magnéticas + anticuerpo + proteína empleando el imán y se lavaron un total de 3 veces en 1 mL de tampón de lisis T frío, con inhibidores de proteasas recién añadidos, durante 5 minutos a 4ºC bajo agitación en rotación. Tras los lavados se suspendieron las muestras en 65 µL de tampón de carga Laemmli 1X (ver a continuación) y se incubaron durante 10 minutos a 95ºC agitándolas por vórtex a mitad de la incubación para su elución. Las inmunoprecipitaciones se almacenaron a -20ºC hasta su uso.

Para evaluar la unión de las proteínas de interés a CCM2 se realizaron western blots (ver sección 4.15). Los anticuerpos y sus condiciones de uso empleados en esta tesis se proporcionan en la **Tabla Suplementaria 5**.

- Tampón de lisis T: 150 mM de NaCl, 10 mM de Tris-HCl pH 7,5, 1% de Triton X-100, 0,5% de Igepal, 1 mM de EDTA pH 8,0 y 1 mM de EGTA pH 8,0 en agua ultrapura. En el momento previo a su uso se añaden inhibidores de proteasas y fosfatasas: 0,2 mM de PMSF, 1X de PIC, 20 mM de NaFl y 20 mM de ortovanadato.
- Tampón de carga Laemmli 1X: 2% de SDS, 10% de glicerol, 1,25% de β-mercaptoetanol, 0,01% de bromofenol azul y 60 mM de Tris-HCl pH 6,8 en agua ultrapura filtrado a través de un filtro de 0,22 µm empleando una jeringuilla.

4.24 DETERMINACIÓN DE LA METILACIÓN GENÓMICA

Para cuantificar los niveles globales de las marcas 5-metilcitosina (5-mC) y 5-hidroximetilcitosina (5-hmC) en el ADN de ESCs WT o KO para *Ccm2* se emplearon los ensayos de inmunoadsorción ligado a enzima (ELISA) 5-mC DNA ELISA y Quest 5-hmC DNA ELISA (ambos de Zymo Research), utilizando 100 ng de ADN genómico y

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

siguiendo las instrucciones del fabricante. La absorbancia, reflejo de la cantidad de 5-mC o 5-hmC en el ADN, se midió a 405-450 nm empleando un lector de placas Mithras LB 940 (Berthol Technologies).

4.25 ENSAYOS DE DIFERENCIACIÓN

4.25.1 Formación de cuerpos embrioides

Las ESCs se diferenciaron a partir de cultivos en suspensión mediante la formación de agregados esféricos denominados cuerpos embrioides (EBs, del inglés *embryoid bodies*), que mimetizan muchas de las características del desarrollo embrionario *in vivo*[251]. Se sembraron ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* a una densidad de 155.000 células/mL en placas Petri de baja adherencia y se cultivaron en medio pluripotente en presencia de suero, pero ausencia de LIF durante 9 días. Se refrescó el medio de cultivo cada 2 días. Para ello, se recogieron los EBs, se precipitaron por gravedad y se aspiró cuidadosamente el sobrenadante. Seguidamente se resuspendieron en medio fresco y se resembraron en placas nuevas. A día 9 se tomaron imágenes representativas para su análisis visual y se extrajo ARN (ver sección 4.11). Las imágenes se obtuvieron empleando un microscopio invertido Olympus IX51 con una cámara Olympus DP72 y se utilizaron para cuantificar el diámetro relativo de los EBs empleando el software Image J (v1.33).

4.25.2 Formación de gastruloides

Las ESCs se diferenciaron a estructuras gastruloides capaces de sufrir procesos similares a la gastrulación y elongación, mimetizando la expresión génica del embrión durante el proceso de gastrulación *in vivo*[252]. Para obtener gastruloides se siguió un protocolo publicado anteriormente[253]. Brevemente, se sembraron 400 ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en placas de 96 pocillos con fondo en U. Se cultivaron durante 120 horas en medio N2B27 compuesto por 50% de medio DMEM/F-12 y un 50% de medio Neurobasal, suplementado con 1X de B-27, 1X de N-2, 2 mM de L-Glutamina, 110 U/mL de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

penicilina/estreptomicina y 9,125 ng/mL de β-mercaptoetanol. Entre las 48 y 72 horas se aplicó un pulso de CHIR99021, suplementando el medio con 3 µM de este inhibidor de GSK3β. Se cambió el medio cada día por medio fresco, con cuidado de no perturbar los agregados celulares formados. Tras 120 horas, se tomaron imágenes representativas para su análisis visual. Las imágenes se obtuvieron empleando un microscopio invertido Olympus IX51 con una cámara Olympus DP72 y se utilizaron para cuantificar la longitud relativa empleando el software Image J (v1.33). Para calcular la eficiencia en la generación de estructuras gastruloides, se comparó el porcentaje de agregados que alcanzaron una morfología similar a la de gastruloide (n=8 o 12 grupos de 12 agregados cada uno).

4.26 ENSAYO DE ESPECIFICACIÓN ENDOTELIAL

Para obtener células progenitoras endoteliales y células endoteliales maduras se trataron placas de cultivo con 18 µg/cm^2 de colágeno IV en PBS y se incubaron al menos 3 horas a 37ºC. A continuación, se sometieron a un tratamiento con luz ultravioleta durante 1 hora, tras lo cual se encuentran listas para la siembra celular. Se sembraron 13.000 células/cm^2 de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en medio N2B27 sin insulina compuesto por 50% de DMEM/F-12 y un 50% de medio Neurobasal, suplementado con 1X de B-27 sin insulina, 1X de N-2, 2 mM de L-Glutamina, 110 U/mL de penicilina/estreptomicina y 9,125 ng/mL de β-mercaptoetanol. Tras 2 días de cultivo, se refrescó el medio añadiendo 3 µM del inhibidor de GSK3β CHIR99021, 4 ng/mL de ACTIVINA A y 12,5 ng/mL de FGFb, y se cultivaron durante 2 días más. Tras estos 2 días de cultivo, se refrescó el medio añadiendo 12,5 ng/mL de FGFb, 20 ng/mL de BMP4, 20 ng/mL de VEGF, 0,25 mM de Br-cAMP y 4 µM del inhibidor de los receptores tipo I ALK4/5/7 de la superfamilia TGFβ (SB431542), y se cultivaron durante 2 días más. A día 4 se obtuvieron células progenitoras endoteliales y se extrajo ARN para su secuenciación masiva (ver sección 4.27). A día 6 se generaron células endoteliales maduras. Se obtuvieron imágenes de ESCs en cultivo y células a día 2, 4 y 6 durante el proceso de especificación

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

empleando un microscopio invertido Olympus IX51 con una cámara Olympus DP72.

4.27 SECUENCIACIÓN MASIVA DE ARN Y ANÁLISIS

El ARN para su secuenciación masiva se extrajo como se describió en la sección 4.11. La compañía responsable de la secuenciación (Novogene) analizó la calidad del ARN mediante el Agilent Bioanalyzer 2100, lo sometió a un protocolo de selección mediante el kit RIboZero para eliminar ARNr (ARN ribosómico) y preparó las librerías. La secuenciación masiva fue realizada usando la plataforma Illumina HiSeq4000, y se generaron lecturas de extremos emparejados de una longitud de 150 pares de bases. Replicados biológicos de ESCs o EPCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* se analizaron mediante secuenciación.

El análisis bioinformático de las muestras de secuenciación masiva se realizó en el laboratorio siguiendo protocolos ampliamente establecidos en el campo. Brevemente, la calidad de las lecturas se analizó usando el software Fastqc (v0.11.5). Las lecturas de baja calidad (puntuación de calidad de la lectura inferior a 30) y los adaptadores de secuenciación se eliminaron usando el software Trimmomatic (v0.39). Seguidamente, las lecturas de alta calidad se mapearon al genoma de ratón (Ensembl GRCm39, versión 104) usando el software de mapeado STAR (v2.7.6) con los parámetros por defecto. Posteriormente, los duplicados de secuenciación se marcaron y eliminaron usando el software de procesado Picard-tools (v2.27.2) mediante los comandos "MarkDuplicates" y "RemoveDuplicates=TRUE". A continuación, las lecturas filtradas se asignaron a los genes correspondientes usando el software htseq-count (v0.13.5), definiendo el parámetro de superposición como "intersection-nonempty". Finalmente, el análisis diferencial de la expresión génica se calculó usando el paquete DESeq2 (v1.32.0), instalado en el ambiente R (v4.2.2). Así, la matriz de contajes de lecturas en genes se cargó mediante el paquete ddsHTseq. Después, se aplicó un pre-filtrado donde se excluyeron del análisis todos los

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

genes que no se expresan mediante la eliminación de las filas en las cuales el sumatorio de las lecturas en todas las muestras para un transcrito es menor a 1 [rowSums(counts(dds)) >= 1]. A continuación, se realizó el análisis diferencial de las comparaciones requeridas, definiendo la muestra control mediante el comando "contrast". Finalmente, las matrices de datos se anotaron añadiendo una lista de atributos útiles para el futuro procesamiento bioinformático (ID de los genes, cromosoma, posición de inicio y fin, hebra, y tipo de transcrito) usando el paquete BiomaRT (v2.52.0) empleando las listas de ratón de Ensembl.

Los datos crudos de RNA-seq publicados anteriormente disponibles en NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) y utilizados a lo largo de este estudio fueron reanalizados siguiendo el mismo análisis bioinformático descrito anteriormente. Los códigos de acceso se proporcionan en la **Tabla Suplementaria 6**.

4.28 ANÁLISIS *IN SILICO* DE LOS DATOS TRANSCRIPTÓMICOS

4.28.1 Análisis de componentes principales

Para comparar los transcriptomas entre diferentes condiciones celulares a partir de los datos de RNA-seq se realizó un análisis de componentes principales (PCA, del inglés *principal component analysis*). Los datos transcripcionales (*counts* normalizadas) se importaron al software Cluster 3.0, se ajustaron los datos centrando los genes a la media (*center genes to mean*), se normalizaron los datos (*normalize genes*) y se aplicó el análisis de los componentes principales a matrices (*apply PCA to arrays*). Los resultados se visualizaron a través del software Origin (Origin Pro 2017b v9.4.0.220).

4.28.2 Análisis mediante GO y KEGG

Para determinar los procesos y vías afectados por la alteración transcripcional causada por la ausencia de CCM1, CCM2 o CCM3 se

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

realizaron análisis de ontología génica (GO, del inglés *gene ontology*) y análisis de vías de señalización KEGG (del inglés *kyoto encyclopedia of genes and genomes*) a partir de los datos obtenidos por RNA-seq. Se empleó la herramienta de anotación funcional DAVID (del inglés *database for annotation, visualization, and integrated discovery*; https://david.ncifcrf.gov)[254] y/o la herramienta funcional de redes de asociación entre proteínas STRING (https://string-db.org/). Se importaron las distintas listas de genes desregulados indicadas para cada caso a lo largo de la sección de resultados de esta tesis. Las listas de genes importadas se analizaron contra una lista de referencia que incluye todos los genes de *Mus musculus* de NCBI. Los códigos GO de los términos afectados o empleados a lo largo de este trabajo se proporcionan en la **Tabla Suplementaria 7**.

4.28.3 Análisis de enriquecimiento GSEA

Para analizar las diferencias de enriquecimiento de la expresión transcripcional para un conjunto de genes entre distintas condiciones se empleó el software GSEA (del inglés *gene set enrichment analysis*; v4.2.2). Se importaron los datos de expresión del RNA-seq (log_2 *counts* normalizadas) de las distintas condiciones en formato de texto delimitado por tabulaciones (txt) y se aportaron las listas de genes deseadas. A continuación, se ejecutó el software con los siguientes ajustes: número de permutaciones igual a 1.000, tipo de permutación "gene_set", estadístico de enriquecimiento "weighted" y se ordenaron los genes por "Diff_of_classes". Los gráficos de enriquecimiento, el valor de enriquecimiento normalizado (NES, del inglés *normalized enrichment score*) y la tasa de descubrimientos falsos (FDR, del inglés *false discovery rate*) se generaron y calcularon a través del programa.

4.28.4 Análisis ChEA3

Para identificar los factores transcripcionales (FT) responsables de los cambios observados en la expresión génica tras la eliminación de CCM2 en ESCs se empleó la herramienta de análisis de enriquecimiento de factores de transcripción (ChEA3, del inglés *ChIP-*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

X enrichment analysis 3; https://maayanlab.cloud/chea3/), que utiliza una biblioteca de factores de transcripción y reguladores epigenéticos y sus presuntos genes diana asociados seleccionados de diversos experimentos basados en técnicas de inmunoprecipitación de la cromatina publicados previamente[255]. Se importó la lista de genes diferencialmente expresados en las ESCs KO para *Ccm2* determinados a través de los datos del RNA-seq.

4.28.5 Búsqueda de motivos consenso

Para determinar los motivos consenso existentes en los promotores de los genes diferencialmente expresados comunes en células progenitoras endoteliales tras la eliminación de *Ccm1*, *Ccm2* y *Ccm3* se empleó el conjunto de herramientas HOMER (v4.11). Concretamente, se utilizó la herramienta findMotifs.pl para buscar motivos consenso de longitudes entre 6 y 12 pares de bases en una ventana de 4 kb alrededor del sitio de inicio de transcripción (TSS) entre los genes de interés.

4.28.6 Análisis de firmas transcriptómicas alteradas por fármacos y pequeñas moléculas

La búsqueda de fármacos o moléculas químicas que puedan provocar firmas génicas opuestas a la generada por la ausencia de CCM1, CCM2 y CCM3 en células progenitoras endoteliales se realizó mediante la herramienta online L1000FWD (https://maayanlab.cloud/L1000FWD/)[256]. Se importaron las listas individuales de genes reprimidos y aumentados determinados a través de los datos del RNA-seq, obteniendo una puntuación de similitud entre las firmas génicas (negativa-opuesta; positiva-mímica) para todas las moléculas contenidas en esta base de datos.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.29 INMUNOPRECIPITACIÓN DE CROMATINA SEGUIDA DE SECUENCIACIÓN MASIVA

4.29.1 Inmunoprecipitación de cromatina y purificación del ADN

La inmunoprecipitación de cromatina (ChIP) unida a ^{3xFL}CCM2, H3K4me1 o H3K4me3 se realizó como se ha descrito en la literatura con pequeñas modificaciones[257]. Se prepararon alícuotas de $10x10^6$ células ESCs WT, *Ccm2* KO o *Ccm2* KO +^{3xFL}CCM2, se centrifugaron 5 minutos a 300 xg y se resuspendieron en PBS para su lavado. A continuación, se recogieron otra vez mediante centrifugación y se resuspendieron las células sedimentadas en 10 mL de PBS. Se añadió un volumen 1/10 (1 mL) de solución de formaldehido 11% recién elaborada (ver a continuación) para su fijación (*crosslink*) y se incubaron durante 10 minutos a RT bajo agitación suave (40-60 rpm). Rápidamente se añadió un volumen 1/20 (577 µL) de glicina 2,5 M para detener la acción del formaldehido y se incubó la solución celular 5 minutos a RT bajo agitación suave (40-60 rpm). Se recogieron las células centrifugando durante 5 minutos a 700 xg a 4ºC y se lavaron dos veces con PBS repitiendo la centrifugación. A partir de este momento las células se mantuvieron siempre en hielo. Se descartó el sobrenadante y se resuspendieron las células sedimentadas en 5 mL de tampón de lisis 1 conteniendo inhibidores de proteasas recién añadidos (ver a continuación). Se incubaron durante 10 minutos a 4ºC en agitación, se centrifugaron durante 5 minutos a 1.350 xg a 4ºC y se descartó el sobrenadante. Se resuspendieron las células sedimentadas en 5 mL de tampón de lisis 2 conteniendo inhibidores de proteasas recién añadidos (ver a continuación) y se incubaron durante 10 minutos a RT bajo agitación en rotación. Se centrifugaron durante 5 minutos a 1.350 xg a 4ºC y se descartó el sobrenadante. Se resuspendieron las células sedimentadas en 3 mL de tampón de lisis 3 conteniendo inhibidores de proteasas recién añadidos (ver a continuación). Se sonicaron las muestras empleando el sonicador digital Sonifier de Branson aplicando 12 ciclos (1 ciclo: 30 segundos ON y 2 minutos OFF) con una amplitud

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

del 20% manteniendo siempre las muestras en un baño de agua-hielo. Seguidamente, se añadió un volumen 1/10 (300 µL) de Triton X-100 10% y se centrifugaron las muestras durante 10 minutos a 18.000 xg a 4ºC. Los lisados (cromatina) se almacenaron a -80ºC hasta su uso.

Para la inmuprecipitación (IP) se trataron tubos de baja adherencia DNA LoBind con solución de bloqueo (ver a continuación) y se añadió lisado con inhibidores de proteasas recién añadidos. Se separó el 10% de la mezcla como *input* y se añadió a la cantidad restante 1 µg de anticuerpo contra FLAG o 2 µg de anticuerpo contra H3K4me1 o H3K4me3 cada 10 µg de cromatina. Se incubaron durante toda la noche a 4ºC bajo agitación en rotación. Al día siguiente se prepararon bolas magnéticas (Dynabeads Protein G) a 4ºC lavándolas tres veces en 1 mL de solución de bloqueo fría en un tubo de baja adherencia. Se recogieron las bolas empleando el imán DynaMag-2. Se resuspendieron 225 ng de bolas por µg de anticuerpo utilizado en 200 µL de solución de bloqueo y se añadieron a la mezcla de cromatina + anticuerpo. Se incubaron durante 4 horas a 4ºC bajo agitación en rotación. Se recogieron las mezclas cromatina + anticuerpo + bolas magnéticas empleando el imán y se resuspendieron en 1 mL de tampón de lavado bajo en sales (ver a continuación). Se incubaron durante 10 minutos a 4ºC bajo agitación en rotación y se recogieron con el imán. Se repitió el mismo proceso para los lavados con 1 mL de tampón de lavado alto en sales, tampón de LiCl y tampón TE (ver a continuación). Se recogieron las muestras con el imán, se resuspendieron en 200 µL de tampón de elución (ver a continuación) y se incubaron durante 15 minutos a 65ºC agitando mediante vórtex cada 2 minutos para separar la cromatina de las bolas magnéticas. Finalmente, se centrifugaron las muestras durante un minuto a 16.000 xg a RT y se procedió a revertir el *crosslink*.

Se revirtió el *crosslink* mediante incubación durante toda la noche a 65ºC. Se trató también el *input* reservado, añadiendo 3 volúmenes de tampón de elución e incubando igualmente a 65ºC. Se recogió el sobrenadante empleando el imán DynaMag-2 y se añadieron 200 µL de tampón TE con el fin de diluir la concentración de SDS

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

presente en el tampón de elución para permitir la purificación del ADN. Se añadió ribonucleasa A a una concentración final de 0,2 mg/mL, se mezcló por inversión y se incubó durante 2 horas a 37ºC. A continuación, se añadió proteinasa K a una concentración final de 0,2 mg/mL, se mezcló por inversión y se incubó durante 2 horas a 55ºC. Se añadieron 400 µL de fenol:cloroformo:isoamil alcohol 25:24:1, se mezclaron las muestras mediante agitación por vórtex y se separó el ADN empleando tubos 5PRIME Phase Lock Gel Heavy siguiendo las instrucciones del fabricante. Se recogió la fase acuosa conteniendo el ADN y se le añadió NaCl a una concentración final de 200 mM, un total de 30 µg de glucógeno y 800 µL de etanol. Se incubaron las mezclas durante 30 minutos a -20ºC y se centrifugaron durante 5 minutos a 18.000 xg a 4ºC. Se lavó el ADN sedimentado añadiendo 500 µL de etanol 80% y se volvió a recoger empleando las mismas condiciones de centrifugación. Se desechó el sobrenadante y se dejó secar al ADN al aire sin llegar a perder la apariencia húmeda. Se resuspendió cada muestra de ADN en 70 µL de 10 mM Tris-HCl pH 8,0 (ver a continuación). Tras la purificación se analizó la calidad y cantidad de ADN empleando un espectrofotómetro Nanodrop 2000. Los anticuerpos y sus condiciones de uso empleados en esta tesis se proporcionan en la **Tabla Suplementaria 5**.

- Solución de formaldehido 11%: 50 mM de HEPES-KOH pH 7,3, 100 mM de NaCl, 1 mM de EDTA, 0,5 mM de EGTA y 11% de formaldehido en agua ultrapura.

- Tampón de lisis 1: 50 mM de HEPES-KOH pH 7,3, 140 mM de NaCl, 1 mM de EDTA, 10% de glicerol, 0,5% de Igepal y 0,25% de Triton X-100 en agua ultrapura. Añadir frescos los inhibidores de proteasas PMSF y PIC.

- Tampón de lisis 2: 10 mM de Tris-HCl pH 8,0, 200 mM de NaCl, 1 mM de EDTA y 0,5 mM de EGTA en agua ultrapura. Añadir frescos los inhibidores de proteasas PMSF y PIC.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

- Tampón de lisis 3: 10 mM de Tris-HCl pH 8,0, 100 mM de NaCl, 1 mM de EDTA, 0,5 mM de EGTA, 0,1% de desoxicolato sódico y 0,5% de N-lauroilsarcosina en agua ultrapura. Añadir frescos los inhibidores de proteasas PMSF y PIC.
- Solución de bloqueo: 0,05% de BSA en PBS.
- Tampón de lavado bajo en sales: 20 mM de Tris-HCl pH 8,0, 150 mM de NaCl, 2 mM de EDTA, 0,1% de SDS y 1% de Triton X-100 en agua ultrapura.
- Tampón de lavado alto en sales: 20 mM de Tris-HCl pH 8,0, 500 mM de NaCl, 2 mM de EDTA, 0,1% de SDS y 1% de Triton X-100 en agua ultrapura.
- Tampón de LiCl: 20 mM de Tris-HCl pH 8,0, 250 mM de LiCl, 1 mM de EDTA, 1% de desoxicolato sódico y 1% de Igepal en agua ultrapura.
- Tampón TE: 10 mM de Tris-HCl pH 8,0 y 1 mM de EDTA en agua ultrapura.
- Tampón de elución: 50 mM de Tris-HCl pH 8,0, 10 mM de EDTA y 1% de SDS en agua ultrapura.
- 10 mM Tris-HCl pH 8,0 en agua ultrapura filtrado a través de un filtro de 0,22 µm empleando una jeringuilla.

4.28.2 Secuenciación y análisis

La unidad encargada de la secuenciación (CiC bioGUNE) preparó las librerías y analizó la calidad mediante el Agilent Bioanalyzer 2100. La secuenciación se realizó empleando la plataforma Illumina NovaSeq6000, y se generaron lecturas de extremos emparejados de una longitud de 100 pares de bases. El ADN unido a ^{3xFL}CCM2 en ESCs *Ccm2* KO +^{3xFL}CCM2 y su *input* o unido a H3K4me1 o H3K4me3 en ESCs WT o KO para *Ccm2* y sus respectivos

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

inputs se analizaron mediante secuenciación de ADN unido a la cromatina inmunoprecipitada (ChIP-seq).

El análisis bioinformático de las muestras de secuenciación se realizó en el laboratorio siguiendo protocolos ampliamente establecidos en el campo. Brevemente, la calidad de las lecturas se analizó empleando el software Fastqc (v0.11.5). Las lecturas de baja calidad (puntuación de calidad de la lectura inferior a 28) y los adaptadores de secuenciación se eliminaron empleando el software Trimgalore (v0.6.7). Seguidamente, las lecturas de alta calidad se mapearon al genoma de ratón (Ensembl GRCm39, versión 104) empleando el software de mapeado Bowtie2 (v2.4.5) con los parámetros -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 y -X 2000. Las lecturas mapeadas en formato SAM se convirtieron a formato BAM, se unieron los resultados de las réplicas biológicas, en caso de existir, y se ordenaron las secuencias por coordenadas empleando el paquete de herramientas Samtools (v1.15.1). Posteriormente, los duplicados de secuenciación se marcaron y eliminaron usando el software de procesado Picard-tools (v2.27.2) mediante los comandos "MarkDuplicates" y "RemoveDuplicates=TRUE". A continuación, las lecturas filtradas se asignaron a los correspondientes picos de unión en el genoma empleando la herramienta MACS3 (v3.0.07a), definiendo el parámetro --broad en el caso de analizar marcas de histonas. La anotación de los picos para determinar la identidad de regiones conocidas a través de su posición en el cromosoma se realizó mediante el conjunto de herramientas HOMER (v4.11).

Los datos crudos de ChIP-seq publicados anteriormente disponibles en NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) y utilizados a lo largo de este estudio fueron reanalizados siguiendo el mismo análisis bioinformático descrito anteriormente. Los códigos de acceso se proporcionan en la **Tabla Suplementaria 6**.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.30 ANÁLISIS DE PERFILES DE ENRIQUECIMIENTO A PARTIR DE CHIP-SEQ

El perfil global de cada factor de transcripción o marca de histonas en las regiones de interés se realizó empleando el software EaSeq (v1.111)[258]. Se importaron los archivos en formato BAM generados durante el análisis de la secuenciación (ver sección 4.29.2). Las regiones se establecieron importando las coordenadas de los genes diferencialmente expresados tras la eliminación de *Ccm2*. Los perfiles globales se representaron a través de mapas de calor y gráficos de densidad media en torno al sitio de inicio de la transcripción (TSS) en ventanas de 5, 6 o 10 kb (especificadas en la figura correspondiente para cada experimento).

4.31 ANÁLISIS DE LAS PROTEÍNAS QUE INTERACTÚAN CON CCM2

Para determinar las proteínas que se unen a CCM2 en células pluripotentes se realizó una inmunoprecipitación de proteína total de ESCs *Ccm2* KO $+^{3xFL}$CCM2 seguida de un análisis de espectrometría de masas en tándem por cromatografía de nanolíquidos (nLC-MS/MS) siguiendo el protocolo descrito a continuación. Se crecieron 3 placas de 150 mm de diámetro hasta alcanzar una confluencia del 80-90%. Se recogieron las células, se centrifugaron durante 5 minutos a 300 xg y se lavaron con PBS frío. Seguidamente, se recogieron otra vez mediante centrifugación, se calculó el volumen de las células sedimentadas (PCV, del inglés *packed cell volumen*) y se resuspendieron en 5 volúmenes de tampón de lisis con inhibidores de proteasas recién añadidos (ver a continuación). Se incubaron durante 10 minutos a 4ºC y se sonicaron las muestras empleando el sonicador digital Sonifier de Branson aplicando 5 ciclos (1 ciclo: 1 segundo ON y 10 segundos OFF) con una amplitud del 30% manteniendo siempre las muestras en un baño de agua-hielo. Se dejaron reposar durante 10 minutos en hielo y se centrifugaron durante 10 minutos a 10.000 xg a 4ºC para sedimentar cualquier resto celular. Se recogieron los sobrenadantes y se prepararon 100 volúmenes de tampón de diálisis (ver a continuación). Se dializaron

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

las muestras durante 5 horas a 4ºC con agitación constante empleando casetes de diálisis Slide-A-Lyzer G2, 10K MWCO sumergidos en tampón de diálisis. Seguidamente, se recogieron los dializados y se centrifugaron durante 20 minutos a 18.000 xg a 4ºC para sedimentar cualquier resto indeseado. Se guardó el sobrenadante conteniendo los lisados totales y se cuantificó la proteína empleando el kit de ensayo de proteínas BCA Pierce siguiendo las indicaciones del fabricante. Para eliminar posibles uniones inespecíficas se llevó a cabo un pre-aclarado de la muestra añadiendo 150 µg de bolas magnéticas (Dynabeads Protein G) por cada mg de proteína e incubando las mezclas durante 1 hora y 30 minutos a 4ºC bajo agitación en rotación. Se recogieron las bolas empleando el imán DynaMag-2 y se almacenaron las muestras a -80ºC hasta su uso.

Para la inmunoprecipitación (IP) se usaron 25 mg de proteína por reacción de IP, se separó el 10% de la mezcla como *input*, se añadieron a la cantidad restante 2 µg por mg de proteína de anticuerpo primario contra FLAG o IgG como control y se incubaron durante toda la noche a 4ºC bajo agitación en rotación. Al día siguiente, se prepararon 300 µg de bolas magnéticas en tampón de IP (ver a continuación) y se realizaron 2 lavados con el mismo tampón con ayuda del imán. Se resuspendieron las bolas en 200 µL de tampón de IP por reacción, se añadieron a las mezclas de proteína + anticuerpo y se incubaron durante 3 horas a 4ºC bajo agitación en rotación. Seguidamente, se recogieron las bolas magnéticas + anticuerpo + proteína empleando el imán y se lavaron un total de 3 veces en 10 mL en tampón de IP con inhibidores de proteasas recién añadidos durante 10 minutos a 4ºC bajo agitación en rotación. Tras los lavados se resuspendieron las muestras en 500 µL de tampón de IP y se incubaron durante otros 10 minutos a 4ºC bajo agitación en rotación. Posteriormente, se recogieron las bolas magnéticas + anticuerpo + proteína empleando el imán y se resuspendieron en 200 µL de tampón de carga Laemmli 2X para su elución (ver a continuación) incubando las muestras durante 5 minutos a 95ºC agitándolas por vórtex a mitad

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de la incubación. Se recogieron las bolas empleando el imán, se recogió el sobrenadante y se repitió el proceso añadiendo otros 200 µL de tampón de carga Laemmli 2X a las bolas. Se recogió de nuevo el sobrenadante y se realizó un ciclo final de elución con 100 µL de tampón. Se recogió este tercer sobrenadante, obteniendo un total de 500 µL. Se conservaron las muestras a -80ºC hasta su uso. Los anticuerpos y sus condiciones de uso empleados en esta tesis se proporcionan en la **Tabla Suplementaria 5**.

Para el envío de las muestras a las instalaciones externas encargadas del análisis nLC-MS/MS (CiC bioGUNE), se concentraron los 500 µL de muestra empleando filtros Amicon Ultra de 0,5 mL y 3kDa siguiendo las instrucciones del fabricante. Tras la concentración se obtuvieron aproximadamente 50 µL de muestra, que se resolvió entonces mediante electroforesis a 100 voltios en un gel de Tris-Glicina con gradiente de 4-20% hasta que las muestras se separaron aproximadamente 3 cm de los pocillos. Finalmente, el gel se selló y se envió para su análisis.

- Tampón de lisis: 25 mM de Tris-HCl pH 7,5, 20% de glicerol, 100 mM de KCl, 0,5 mM de EDTA y 10% de Igepal en agua ultrapura filtrado a través de un filtro de 0,22 µm empleando una jeringuilla.
- Tampón de diálisis: 20 mM de HEPES pH 7,9, 25% de glicerol, 1,5 mM de $MgCl_2$ y 0,2 mM de EDTA en agua ultrapura filtrado.
- Tampón de IP: 20 mM de HEPES pH 7,9, 25% de glicerol, 1,5 mM de $MgCl_2$, 0,2 mM de EDTA y 0,02% de Igepal en agua ultrapura filtrado.
- Tampón de carga Laemmli 2X: 4% de SDS, 20% de glicerol, 10% de β-mercaptoetanol, 0,02% de bromofenol azul y 120 mM de Tris-HCl pH 6,8 en agua ultrapura filtrado.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.32 ANÁLISIS DE REDES DE INTERACCIÓN

Para visualizar y analizar las redes de interacción entre proteínas se utilizó el software *Cytoscape* (v3.8.2)[259]. Se emplearon las aplicaciones de STRING y ClueGO para analizar las interacciones y los procesos biológicos en los que intervienen las proteínas. Se importaron al programa tanto las listas correspondientes de genes desregulados pertenecientes a la cascada MAPK o la lista de interactores de CCM2 como los datos de RNA-seq de cambio transcripcional y significancia estadística [valor de *fold-change* (FC) y FDR, respectivamente]. El color de los círculos viene determinado por el FC mientras que el tamaño viene dado por el valor $-\log_2$ FDR.

4.33 ANÁLISIS ESTADÍSTICOS

Todos los análisis estadísticos se realizaron empleando el software GraphPad Prism 9 o R. La significancia se identificó aplicando t-test de Student de dos colas no pareado o aplicando la prueba ANOVA de un factor para el análisis de la varianza seguida de la prueba de comparaciones múltiples de Dunnett para el cálculo estadístico. Las pruebas empleadas durante esta tesis se indican en la leyenda de las figuras para cada caso. Los *P* valores inferiores a 0,05 se consideraron significativos.

4.34 RECURSOS BIBLIOGRÁFICOS

La revisión bibliográfica de la presente tesis se elaboró mediante las plataformas digitales de libre acceso Google Académico y Pubmed (https://scholar.google.es/ y https://www.ncbi.nlm.nih.gov/pubmed/, respectivamente) para la búsqueda de las referencias utilizadas. La introducción de estas referencias en el manuscrito se realizó a través del software EndNote X9 (https://endnote.com/).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.35 SOFTWARE Y ALGORITMOS

Software	Fuente	Enlace
BiomaRT v2.52.0	Durink *y cols.*, 2009[260]	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
Bowtie2 v2.4.5	Johns Hopkins University	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
ChEA3	Keenan *y cols.*, 2019[255]	https://maayanlab.cloud/chea3/
ChemDraw v22	PerkinElmer Infomatics	https://perkinelmerinformatics.com/products/research/chemdraw
Cluster v3.0	Stanford University	http://bonsai.hgc.jp/~mdehoon/software/cluster/
CRISPOR	Concordet y Haeussler, 2018[243]	http://crispor.tefor.net
Cytoscape v3.8.2	Shannon *y cols.*, 2003[259]	https://cytoscape.org
DAVID	LHRI	https://david.ncifcrf.gov/tools.jsp
DESeq2 v1.32.0	Love *y cols.*, 2014[261]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
EaSeq v1.111	Lerdrup *y cols.*, 2016[258]	http://easeq.net
Fastqc v0.11.5	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc
FlowJo v7.6.1	FlowJo, LLC	https://www.flowjo.com/
Genetic Perturbation Platform (GPP)	Broad Institute	https://portals.broadinstitute.org/gpp/public/
GraphPad Prism 9	GraphPad	https://www.graphpad.com/
GSEA v4.2.2	Broad Institute	https://www.gsea-msigdb.org/gsea
HOMER v4.11	Heinz *y cols.*, 2010[262]	http://homer.ucsd.edu/homer/
htseq-count v0.13.5	Putri *y cols.*, 2022[263]	https://htseq.readthedocs.io/en/master/
Image J v1.33	NIH	https://imagej.nih.gov/ij
L1000FWD	Wang *y cols.*, 2018[256]	https://maayanlab.cloud/L1000FWD/

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Leica LAS X v3.4.2.18368	Leica Microsystems	https://www.leica-microsystems.com
MACS3 v3.0.07a	Zhang *y cols.*, 2008[264]	https://macs3-project.github.io/MACS/
Origin Pro 2017 v9.4.0.220	OriginLab	https://www.originlab.com/Origin
Picard-tools v2.27.2	Broad Institute	https://broadinstitute.github.io/picard/
R v4.2.2	The R Foundation	https://www.r-project.org
Samtools v1.15.1	Li *y cols.*, 2009[265]	https://github.com/samtools/samtools
Sequencher 5.4	Gene Codes Corporation	http://www.genecodes.com/
SerialCloner v2.6.1	Serial Cloner	http://serialbasics.free.fr/Serial_Cloner.html
STAR v2.7.6	Dobin *y cols.*, 2013[266]	https://github.com/alexdobin/STAR
STRING	Szklarczyk *y cols.*, 2021[267]	https://string-db.org/
Trimgalore v0.6.7	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Trimmomatic v0.39	Bolger *y cols.*, 2014[268]	http://www.usadellab.org/cms/?page=trimmomatic
EndNote X9	The EndNote Team	https://endnote.com/

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5. RESULTADOS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.1 ANALIZAR EL PAPEL DE LOS GENES *CCM* EN EL MANTENIMIENTO DE PLURIPOTENCIA

5.1.1 Los genes *Ccm* muestran patrones de expresión específicos

Numerosas evidencias han mostrado el papel clave de los genes *Ccm* en la función endotelial y en los procesos de angiogénesis tanto en la etapa adulta como durante el desarrollo embrionario. Sin embargo, hasta el momento no se conoce su importancia en el contexto pluripotente, en el comienzo de los procesos de diferenciación o en la correcta especificación de linajes y tipos celulares concretos, como por ejemplo en la especificación endotelial. Para definir las posibles funciones que *Ccm1*, *Ccm2* y *Ccm3* puedan ejercer sobre el estado celular pluripotente, lo primero que me propuse fue comprobar y analizar sus niveles de expresión con el fin de averiguar patrones de expresión génica similares entre estos tres genes. Primero, comparé entre varios tejidos la expresión de cada uno de los genes *Ccm* proporcionada en la base de datos "*BioGPS*" (http://biogps.org), portal que centraliza la información de expresión génica distribuida en diferentes fuentes de información[269]. Se observó que cada uno de estos tres genes presentaban un patrón de expresión específico, siendo el bazo el tejido con más expresión de *Ccm1* y *Ccm2*, mientras que la mayor expresión de *Ccm3* se correspondía con hueso, seguido de cerca por intestino delgado, microglía, células troncales embrionarias (ESCs, del inglés *embryonic stem cells*), córtex cerebral y epidermis (**Figura R1A**). Seguidamente, analicé el patrón de expresión de estos genes durante las primeras fases del desarrollo embrionario. Para ello utilicé un set de datos publicado previamente en el que se analiza la expresión mediante secuenciación masiva de ARN de célula única (scRNA-seq) desde la etapa de oocito hasta las células de la masa interna (ICM, del inglés *inner cell mass*), pasando por el estado de 2-células, 4-células y 8-células, células del trofectodermo (TE) y ESCs derivadas de la ICM[270]. Estos datos revelaron una tendencia al alza en la expresión de *Ccm1*, *Ccm2* y *Ccm3*, aunque este último de manera menos

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

pronunciada, a medida que se avanza en el proceso de desarrollo, encontrando los niveles de expresión más bajos en las etapas totipotentes e incrementándose en las pluripotentes (**Figura R1B**). Además, pude observar que, en las ESCs, el gen que más se expresaba era *Ccm2*, seguido por *Ccm1* y *Ccm3* (**Figura R1B**). Finalmente, analicé la expresión de los genes *Ccm* en células somáticas (fibroblastos) y células pluripotentes (ESCs) mediante PCR cuantitativa (qPCR). Se observó que la expresión de cada uno de los genes se encontraba más enriquecida en células pluripotentes que en células somáticas (**Figura R1C**). En conjunto, estos resultados indican que los tres genes *Ccm* están presentes en ESCs, y sugieren un posible papel importante de los mismos durante la etapa de desarrollo embrionario temprano y en el estado celular pluripotente.

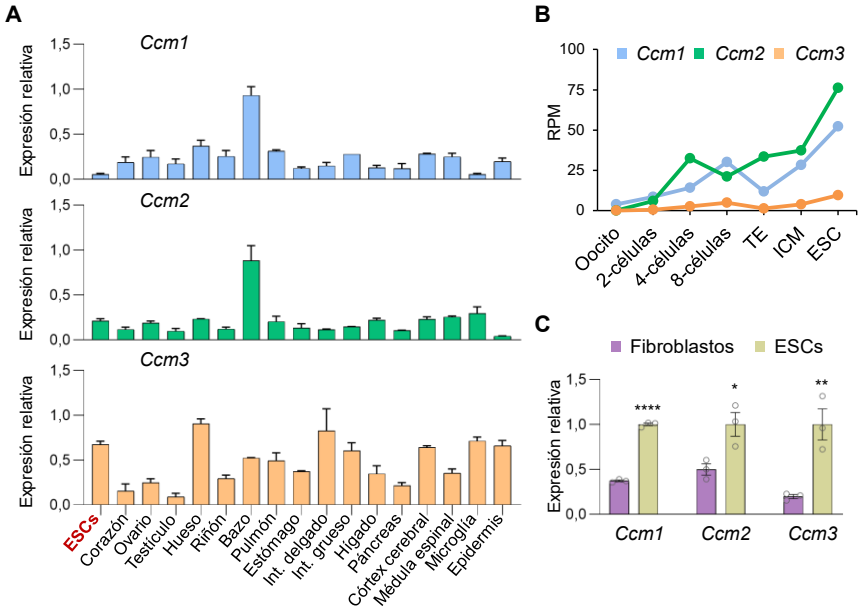


Figura R1. El análisis de los patrones de expresión de los genes *Ccm* revela su presencia en células pluripotentes. **A)** Expresión relativa de *Ccm1* (arriba), *Ccm2* (centro) y *Ccm3* (abajo) en células pluripotentes (ESCs) y diferentes tejidos. Se emplearon los datos de expresión aportados en la base de datos "BioGPS". **B)** Expresión representada en lecturas por millón (del inglés *reads per million*) de los

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

genes *Ccm* durante las etapas del desarrollo embrionario temprano: oocito, 2- 4- y 8-células, trofectodermo (TE) y células de la masa interna (ICM). Se incluyen los datos en ESCs. Se emplearon los datos de scRNA-seq de célula única aportados por Tang y colaboradores (GSE22182)[270]. **C)** Expresión relativa de los genes *Ccm* en células somáticas (fibroblastos) y pluripotentes (ESCs) calculada mediante qPCR. En **A** y **C** los datos se presentan como media ± s.e.m. En **C** n=3 experimentos independientes. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico. *p<0,05; **p<0,01; ****p<0,0001.

5.1.2 Generación de ESCs KO para cada gen *Ccm*

Una vez comprobado que los genes *Ccm* se expresan en el contexto de pluripotencia, decidí analizar su relevancia funcional en la regulación del estado celular pluripotente. Para este fin, realicé una aproximación de pérdida de función generando líneas ESCs *knock-out* (por simplicidad, ESCs KO) de origen murino (cepa 129S4/SvJae) para cada uno de los genes mediante el sistema de edición génica CRISPR-Cas9[271,272]. La estrategia empleada para la obtención de las ESCs KO se representa en la **Figura R2A**. Brevemente, diseñé una guía de ARN capaz de dirigir a la endonucleasa Cas9 específicamente al exón más cercano al codón de inicio de la transcripción de cada gen *Ccm* (**Figuras R2B**, **R2D** y **R2F**). A continuación, transfecté ESCs *wildtype* (por simplicidad, ESCs WT) con un plásmido que contiene la endonucleasa, el ARN guía y un gen de resistencia al antibiótico puromicina, y seleccioné durante dos días aquellas células que incorporaron el plásmido. Una vez terminada la selección transitoria, diluí cada cultivo celular con el fin de obtener clones independientes de cada intervención génica realizada, que fueron inicialmente cribados para homocigosis con deleción en cada condición mediante una estrategia de genotipado por PCR (datos no mostrados). De cada población clonal candidata extraje ADN, ARN y proteína para validar las líneas KO generadas. El análisis por secuenciación Sanger del ADN muestra las deleciones generadas, lo que me permitió simular la nueva secuencia proteica modificada y detectar alteraciones en la secuencia de aminoácidos y sitios de parada de la traducción prematuros (**Figuras R2B**, **R2D** y **R2F**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

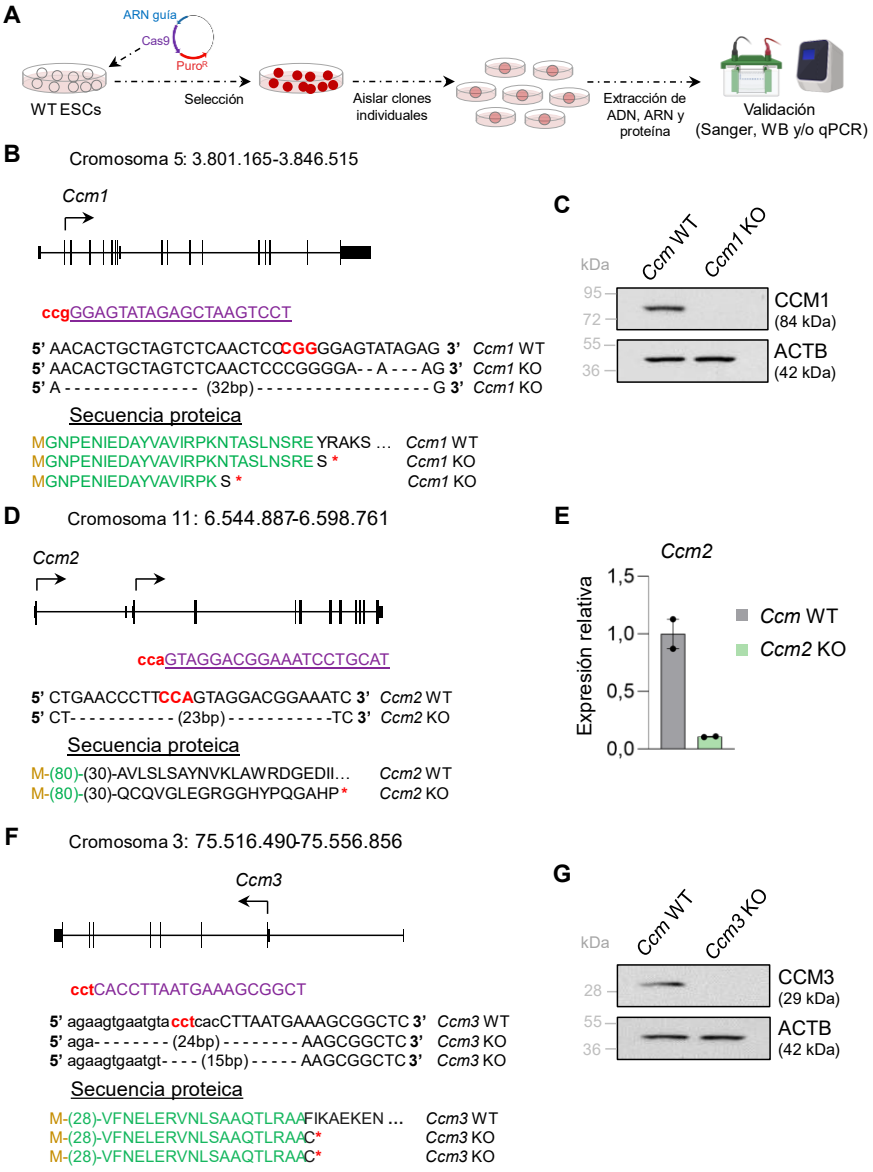


Figura R2. Generación de líneas KO para *Ccm1*, *Ccm2* y *Ccm3* en ESCs. A) Representación esquemática de la estrategia seguida para generar líneas KO empleando la técnica CRISPR-Cas9. Se transfectaron ESCs WT con un plásmido que contiene la guía de ARN, la Cas9 y un gen de resistencia a puromicina. Se seleccionaron las células que incorporaron el plásmido con antibiótico durante 2 días.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Se aislaron clones individuales y se extrajo ADN, ARN y proteína para su validación mediante secuenciación Sanger y WB o qPCR. **B)** Representación esquemática de *Ccm1* indicando el codón de inicio de la transcripción (ATG) con una flecha. Se indica el exón diana y la secuencia de la guía de ARN (violeta) con la secuencia PAM (rojo). La secuenciación Sanger permitió caracterizar la mutación generada, consistente en la eliminación de 5 pares de bases (bp) en un alelo y 32 bp en el otro. La simulación de la secuencia proteica resultante permite identificar alteraciones en la secuencia y sitios de parada de la traducción prematuros (asterisco). **C)** Validación de la pérdida de CCM1 mediante western blot. ACTB se empleó como control de carga. **D)** Representación esquemática de *Ccm2* igual que en **B**. La secuenciación Sanger permitió identificar una deleción de 23 bp generando alteraciones en la secuencia proteica y un sitio de parada de la traducción prematuro (asterisco). **E)** Validación de la pérdida de *Ccm2* mediante qPCR. Los datos se presentan como media ± s.e.m. **F)** Representación esquemática de *Ccm3* igual que en **B**. La secuenciación Sanger permitió identificar una deleción de 24 bp en un alelo y 15 bp en el otro (8 bp correspondientes al exón en ambos casos) generando alteraciones en la secuencia proteica y un sitio de parada de la traducción prematuro (asterisco). **G)** Validación de la pérdida de CCM3 mediante western blot. ACTB se empleó como control de carga. En **B**, **D** y **F** se representa la metionina inicial (amarillo), la secuencia proteica invariable tras las mutaciones (verde), la secuencia que varía (negro) y entre paréntesis el número de aminoácidos correspondiente a cada sección. Imagen creada con BioRender.com.

El análisis mediante western blot y qPCR confirmó la ausencia de las proteínas CCM1 y CCM3, y una pérdida prácticamente total de los niveles de ARN de *Ccm2*, respectivamente (**Figuras R2C**, **R2E** y **R2G**). La expresión de *Ccm2* se analizó mediante qPCR debido a que no disponía de un anticuerpo comercial que funcionase para su detección en células de origen murino.

5.1.3 La función CCM es dispensable en el mantenimiento de pluripotencia

Para establecer si la falta de función de cada proteína CCM afecta al mantenimiento del estado pluripotente, llevé a cabo varias aproximaciones clásicas en el campo. Primero, analicé la morfología y el patrón de tinción de la fosfatasa alcalina (AP, del inglés *alkaline phosphatase*) de las colonias de ESCs WT y KO para *Ccm1*, *Ccm2* y *Ccm3* en condiciones estándar de cultivo, que se corresponden a la presencia de suero y la citoquina LIF (por simplicidad, SL). Mencionar que las células pluripotentes presentan niveles elevados de AP

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

comparado con la mayoría de células somáticas, por lo que la detección de su actividad se usa rutinariamente como marcador específico del estado celular pluripotente[250]. Se observó que cada población de ESCs KO mostraba una morfología y patrón de tinción de AP similar a las células WT tras varios pases en cultivo (>10) (**Figura R3A**). Cabe destacar que las células sin CCM2 mostraban colonias más compactas y regulares que el resto de las condiciones. Segundo, el análisis mediante western blot e inmunofluorescencia mostraron que la expresión de diferentes marcadores de pluripotencia (OCT4, SSEA-1 y ZFP281) no se veían afectados de modo evidente ni a nivel de cantidad total de proteína ni en su localización subcelular (**Figuras R3B** y **R3C**).

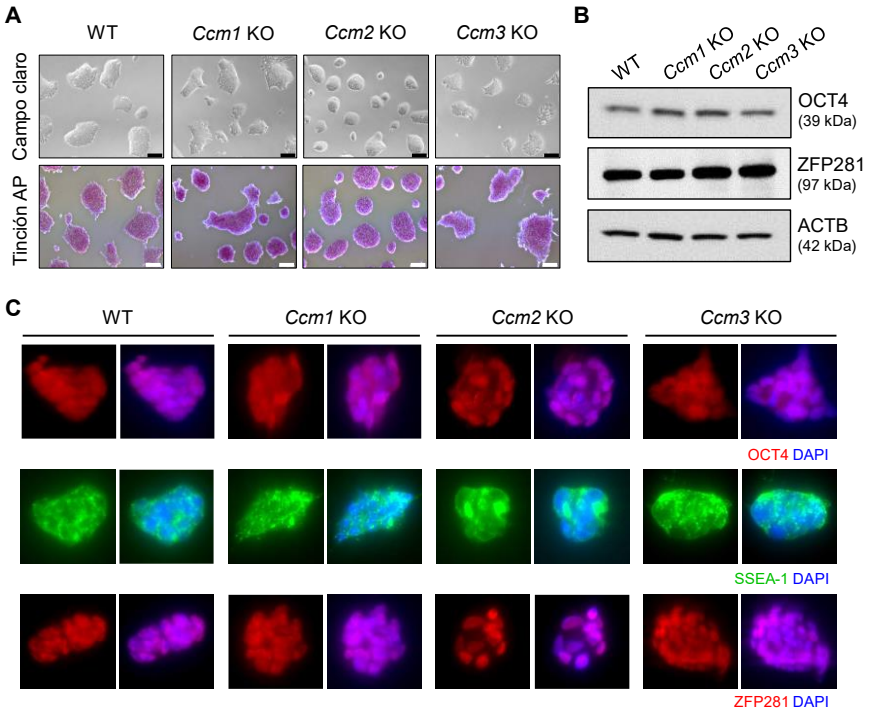


Figura R3. La ausencia de cada gen *Ccm* no conlleva la pérdida de pluripotencia. A) Imágenes representativas en campo claro (arriba) y de tinción de la fosfatasa alcalina (AP, abajo) de las ESCs WT, *Ccm1* KO, *Ccm2* KO y *Ccm3* KO en condiciones estándar de cultivo. Las barras representan 100 µm. **B)** Western blot para detectar los niveles proteicos de los marcadores pluripotentes OCT4 y ZFP281 en cada

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

población de ESCs indicada. ACTB se empleó como control de carga. **C)** Imágenes representativas de los niveles y localización celular de OCT4, SSEA-1 y ZFP281 obtenidas mediante ensayo de inmunofluorescencia en las condiciones indicadas. Los núcleos se marcaron con DAPI.

En tercer lugar, analicé el estado del ciclo celular y la ratio de proliferación celular en las ESCs WT y en cada línea mutante de pérdida de función CCM. Las células pluripotentes se caracterizan por una rápida proliferación y un ciclo celular corto, aproximadamente cada 12 horas se completa un ciclo, que presenta una fase S prolongada en el tiempo en detrimento de las fases G1 y G2[273,274]. Para analizar el ciclo celular, cuantifiqué el contenido de ADN de las células mediante tinción con yoduro de propidio y citometría de flujo activada por fluorescencia (FACS, por sus siglas en inglés). El yoduro de propidio es un tinte fluorescente que se une a las moléculas de ADN de manera proporcional a su cantidad, lo que permite su cuantificación teniendo en cuenta que las células en fase S del ciclo celular duplican su contenido en ADN (2n-4n) con respecto a las células en la fase previa de crecimiento G1 (2n), que se mantendrá duplicado durante la fase G2 (4n) hasta que se reparta entre las células resultantes tras finalizar en la fase M la división celular (2n). Los resultados mostraron, en consonancia con la literatura al respecto de células pluripotentes, una mayor cantidad de células en fase S en todas las condiciones analizadas (**Figura R4A**). Además, los resultados obtenidos revelaron que las ESCs KO para *Ccm1*, *Ccm2* o *Ccm3* presentaban proporciones similares de distribución del ciclo celular a las ESCs WT (**Figura R4A**). Con el fin de analizar la proliferación celular durante un período de seis días, y partiendo del mismo número de células para todas las condiciones, cuantifiqué el número total de éstas cada dos días (**Figura R4B**). En este caso, tampoco se observaron diferencias importantes entre los patrones de proliferación celular de cada una de las ESCs KO con respecto a las ESCs WT (**Figura R4C**).

En conjunto, los resultados obtenidos revelan que la pérdida de función de CCM1, CCM2 o CCM3 es compatible con el mantenimiento

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de un estado celular pluripotente en condiciones estándar de cultivo (SL).

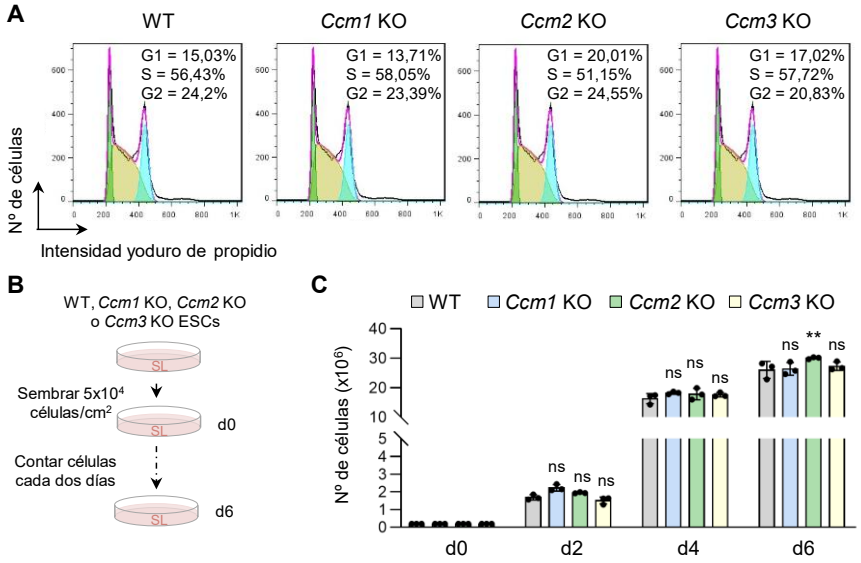


Figura R4. La eliminación de las proteínas CCM no afecta a la proliferación de las células pluripotentes. A) Análisis del ciclo celular por FACS mediante tinción con yoduro de propidio en ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en condiciones estándar de cultivo. Se muestra el porcentaje de células en las fases G1, S y G2 para todas las condiciones. **B)** Representación esquemática de la estrategia seguida para analizar la proliferación celular. Se sembraron 5x10^4 células/cm^2 de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en condiciones estándar de cultivo (SL) y se contó el número total de células cada 2 días durante un total de 6. **C)** Representación del número de células en los días especificados para todas las condiciones. n=3 experimentos independientes. Los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre cada KO y las ESCs WT. **p<0,01; ns: no significativo.

5.1.4 La ausencia de CCM2 favorece la autorrenovación de las ESCs

Tras haber comprobado que la ausencia de cada proteína CCM no conlleva una pérdida del estado celular pluripotente, me propuse analizar posibles efectos de estos genes en la capacidad de autorrenovación de las ESCs. Para este fin, realicé una aproximación

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

conocida en el campo denominada ensayo de formación de colonias. Este ensayo consiste en la siembra de ESCs a baja densidad para evaluar su progreso a nivel de célula única. Para este fin, las ESCs WT y mutantes para cada gen *Ccm* se mantuvieron en cultivo en presencia o ausencia de la citoquina LIF, que resulta indispensable para mantener el estado pluripotente en condiciones de cultivo estándar. Luego, se evaluó su capacidad para proliferar y formar colonias pluripotentes mediante la tinción AP (**Figuras R5A** y **R5B**). Las colonias formadas se clasificaron según su nivel de tinción en: i) colonias no diferenciadas con morfología compacta y tinción AP homogénea, ii) colonias parcialmente diferenciadas con morfología irregular y tinción AP heterogénea y iii) colonias diferenciadas con morfología plana y pérdida completa de tinción AP. Cabe mencionar que el cultivo estándar de ESCs constituye un compendio heterogéneo de estados pluripotentes, por lo que puede producirse la diferenciación espontánea de células dando lugar a colonias parcial o totalmente diferenciadas incluso en presencia de LIF. Por otro lado, la ausencia de LIF favorece la salida del estado celular pluripotente y la consiguiente pérdida de tinción AP. Mis resultados mostraron que en presencia de LIF, las ESCs *Ccm2* KO dieron lugar a un mayor número de colonias no diferenciadas que las WT y las células mutantes para *Ccm1* y *Ccm3* (**Figuras R5C** y **R5D**), sugiriendo que la capacidad de autorrenovación de las ESCs KO para *Ccm2* se mantiene y se encuentra incluso aumentada bajo condiciones estándar de cultivo (SL). Además, en ausencia de LIF se observó que la eliminación de *Ccm2* favoreció la retención del estado celular pluripotente, o al menos retrasó la salida de pluripotencia, al ser capaces estas células de formar colonias no diferencias en comparación con las ESCs WT (**Figuras R5C** y **R5D**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

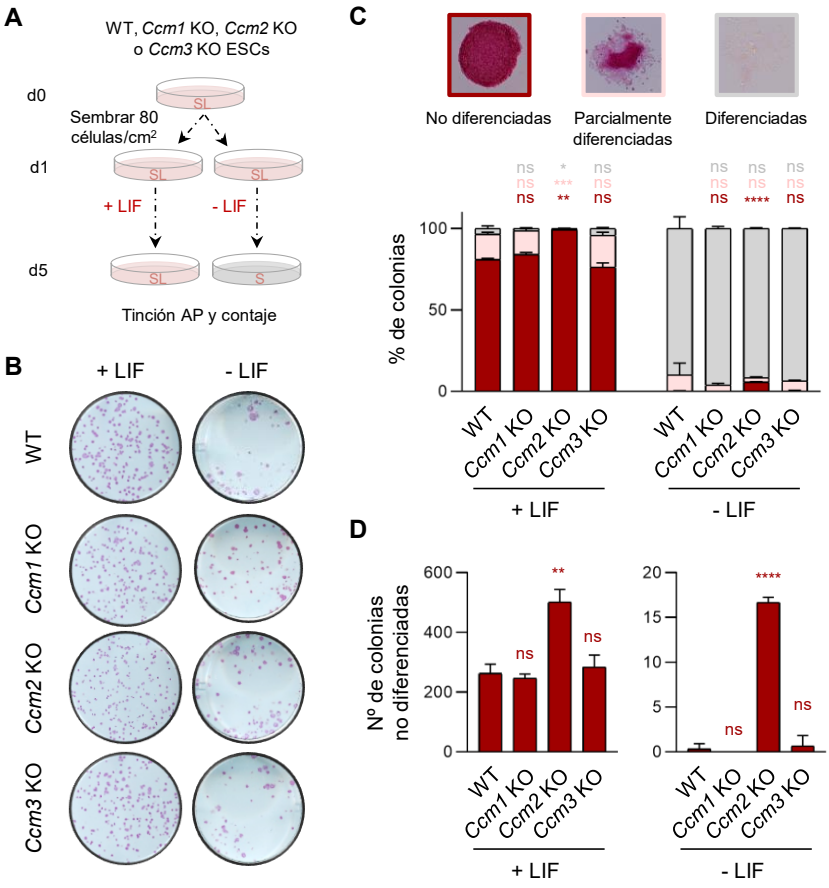


Figura R5. CCM2 afecta negativamente a la capacidad de autorrenovación de las ESCs. A) Representación esquemática de la estrategia seguida para realizar un ensayo de formación de colonias. Se sembraron 80 células/cm^2 de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en condiciones estándar de cultivo (SL). Al día siguiente se cambió el medio por medio pluripotente estándar o medio de diferenciación sin LIF. Tras 5 días se realizó una tinción AP y se contaron las colonias clasificándolas en no diferenciadas, parcialmente diferenciadas y diferenciadas, atendiendo a la presencia y homogeneidad de la tinción. **B)** Imágenes representativas de la tinción AP para todas las condiciones. **C)** (Arriba) Imágenes representativas de la clasificación de las colonias. Una tinción homogénea y morfología compacta se clasificó como colonia no diferenciada, una tinción heterogénea con morfología irregular se clasificó como colonia parcialmente diferenciada y la falta de tinción se clasificó como colonia diferenciada. (Abajo) Representación mediante gráfica de barras del porcentaje de colonias perteneciente a cada clase para todas las condiciones. El color de la barra corresponde al color del marco de las imágenes representativas mostradas arriba. **D)**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Representación mediante gráfica de barras del número total de colonias no diferenciadas para todas las condiciones. En **C** y **D** los datos se corresponden a 3 experimentos independientes y se representan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre cada condición KO y las ESCs WT. *p<0,05; **p<0,01; ***p<0,001; ****p<0,0001; ns: no significativo.

5.1.5 Las CCMs intervienen en el paisaje transcripcional pluripotente

Una vez establecido que las proteínas CCM no son necesarias para mantener el estado pluripotente (**Figuras R3** y **R4**), pero teniendo en cuenta que, al menos en el caso de CCM2, su ausencia parece modificar la capacidad de autorrenovación (**Figura R5**), sumado al hecho de que los patrones de expresión de los tres genes *Ccm* podrían indicar un papel durante la salida de pluripotencia y especificación celular (**Figura R1**), me planteé estudiar los posibles mecanismos moleculares en los que intervienen estas tres proteínas en el estado celular pluripotente. Para ello analicé mediante secuenciación masiva de ARN (RNA-seq, por sus siglas en inglés) el transcriptoma de dos réplicas biológicas de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* cultivadas en condiciones estándar de pluripotencia (SL) (**Figura R6A**). Mediante un análisis de componentes principales (PCA, del inglés *principal component analysis*), se observó que la ausencia de las proteínas CCM causó alteraciones del paisaje transcripcional suficientes para separar cada una de las células mutantes de su contraparte WT (**Figura R6B**). Asimismo, este análisis reveló que las consecuencias en el paisaje transcripcional a nivel global como consecuencia de la eliminación de cada gen *Ccm* es específica a cada condición mutante, puesto que sus componentes principales se agrupaban separadamente entre si (**Figura R6B**). Además, empleando el coeficiente de correlación de Pearson para generar una matriz de similitud para todas las condiciones acoplada a una agrupación jerárquica, se vio que el transcriptoma de las ESCs KO para *Ccm2* representaba un grupo más alejado del resto de condiciones analizadas (**Figura R6C**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

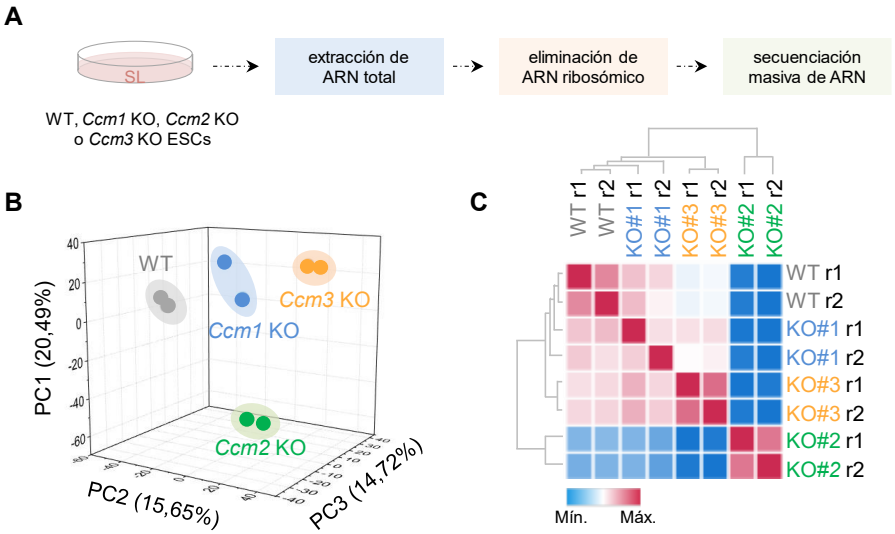


Figura R6. La eliminación de las proteínas CCM altera el perfil transcripcional de las células pluripotentes. A) Representación esquemática de la estrategia seguida para el análisis del transcriptoma. Se sembraron ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en condiciones estándar de cultivo. Se extrajo ARN total mediante TRIZOL, se eliminó el ARN ribosómico mediante el kit RIboZero de Illumina y se procedió a la preparación de las librerías y la secuenciación masiva del ARN (RNA-seq). n=2 experimentos independientes. **B)** Análisis de componentes principales (PCA) de los perfiles transcripcionales de las células ESCs WT y KO para *Ccm1*, *Ccm2* o *Ccm3* empleando los datos del RNA-seq. El porcentaje de variación explicado por cada componente principal (PC) se indica en el eje correspondiente. **C)** Representación mediante mapa de calor del análisis de similitud a partir de los datos del RNA-seq entre los replicados de ESCs WT y KO para *Ccm1*, *Ccm2* o *Ccm3*. Se empleó el coeficiente de correlación de Pearson para generar una matriz de similitud acoplada a una agrupación jerárquica.

Tras observar que la ausencia de las proteínas CCM causa alteraciones transcripcionales específicas a cada una, analicé el transcriptoma de cada condición KO para caracterizar el papel individual de CCM1, CCM2 y CCM3 en los mecanismos moleculares pluripotentes. Con este fin, clasifiqué como transcritos expresados diferencialmente todos aquellos que ven alterada su expresión con cambios iguales o superiores al 50% (FC>1,5) y con una tasa de descubrimientos falsos inferior al 0,05 (FDR <0,05). Así, se identificó

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

que, bajo estos criterios de cambio en la expresión génica, la ausencia de CCM1 provoca que solo 32 genes se vean alterados, mientras que la falta de CCM2 y CCM3 altera la expresión de 1.363 y 418 genes respectivamente (**Figura R7A**). Además, en los tres casos, la gran mayoría de los transcritos afectados (>90%) son ARNs codificantes de proteínas (**Figura R7A**).

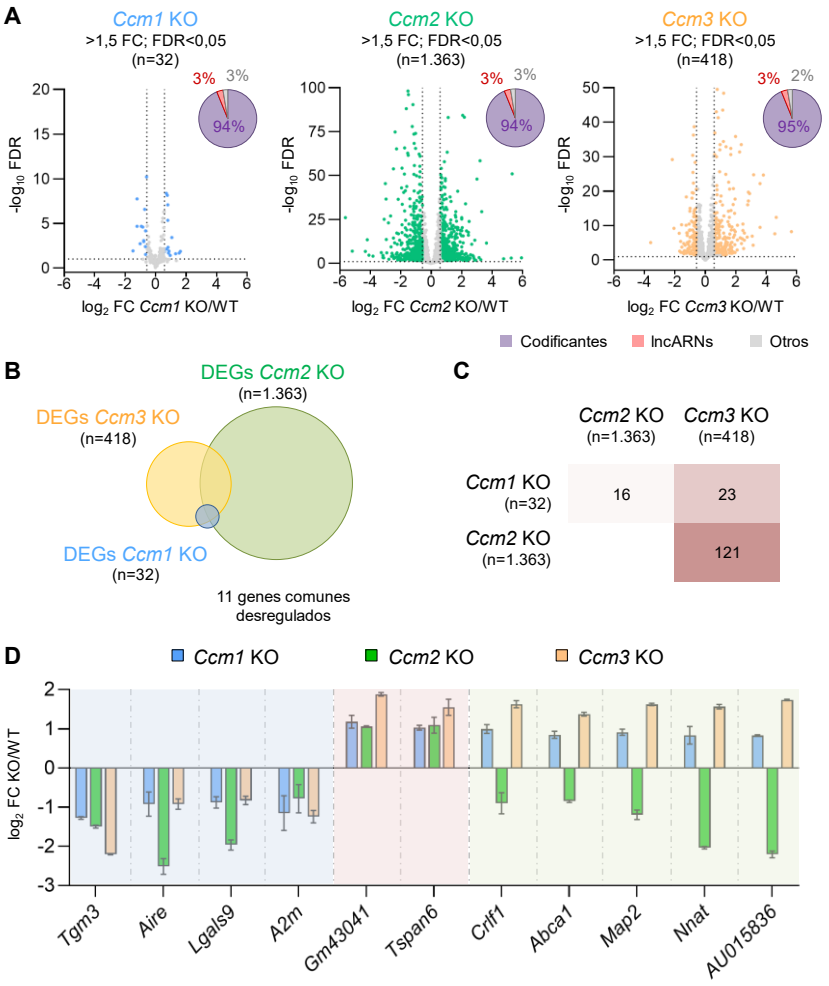


Figura R7. Identificación de los cambios transcripcionales promovidos por la pérdida de función de cada gen *Ccm* en ESCs. A) Representación mediante "*volcano*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

plot" de la expresión génica en *Ccm1* KO (izquierda), *Ccm2* KO (centro) y *Ccm3* KO (derecha) empleando los datos de RNA-seq. Se resaltan los genes diferencialmente expresados (DEGs, del inglés *differential expressed genes*) entre cada KO y el WT (>1,5 FC; FDR<0,05). Se muestra el porcentaje de transcritos desregulados que constituyen ARNs codificantes de proteínas, ARNs largos no codificantes y otros ARNs para cada condición. **B)** Representación mediante diagrama de Venn de los DEGs en las ESCs KO para *Ccm1*, *Ccm2* y *Ccm3* mostrando los genes comunes entre las tres condiciones. **C)** Representación mediante matriz de similitud de los DEGs compartidos entre los pares de condiciones. Se muestra el número de genes desregulados comunes. **D)** Representación mediante gráfica de barras del valor de $\log_2$ FC (*fold-change*) de los 11 genes comunes a *Ccm1* KO, *Ccm2* KO y *Ccm3* KO para todas las condiciones. Los datos se presentan como media ± s.e.m.

Dada la función de los genes *Ccm* en el contexto endotelial y que la pérdida de función independiente de cada uno de ellos provoca la misma enfermedad, decidí estudiar también el papel común que juegan en pluripotencia. Para ello, realicé una comparación a tres con el total de los genes desregulados en cada ESC KO, hallando así solo 11 genes compartidos entre las tres condiciones (**Figura R7B**). Asimismo, determiné el número de genes expresados diferencialmente de manera común entre cada par de condiciones (**Figura R7C**). En consonancia con los resultados extraídos de la matriz de correlación anterior (**Figura R6C**), el patrón de expresión de los 11 genes alterados en los tres KO representó la discordancia en el efecto de la pérdida de CCM2. De esta manera, se observó que seis de los genes mostraron la misma tendencia de desregulación transcripcional en las tres ESCs mutantes (*Tgm3*, *Aire*, *Lgals9* y *A2m* están disminuidos y *Gm43041* y *Tspan6* aumentan con respecto a las ESCs WT), mientras que las ESCs KO para *Ccm2* mostraron una tendencia opuesta para los 5 genes restantes (*Crlf1*, *Abca1*, *Map2*, *Nnat* y *AU015836*) (**Figura R7D**).

El número tan bajo de genes desregulados compartidos entre *Ccm1*, *Ccm2* y *Ccm3* KO impidió un análisis en profundidad de los procesos comunes potencialmente afectados. El hecho de que la ausencia de CCM1 casi no altere el transcriptoma de las ESCs con los criterios escogidos de significancia y cambio en la expresión de al menos un 50% (n=32 genes desregulados con FC>1,5 y FDR <0,05) podría enmascarar efectos que sí pudiesen ser compartidos al menos

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

entre *Ccm2* y *Ccm3* KO. Por esta razón, decidí centrarme en los 121 genes comunes desregulados en ausencia de CCM2 y CCM3 (**Figura R8A**). Analizando la ontología de estos genes mediante el uso de la herramienta DAVID (del inglés *database for annotation, visualization and integrated discovery*) (https://david.ncifcrf.gov), se observó que presentaban un enriquecimiento en procesos biológicos y componentes celulares relacionados con la función endotelial (p. ej. angiogénesis, adhesión, unión celular y membrana entre otros) (**Figura R8B**). Asimismo, y en sintonía con los defectos letales observados durante el desarrollo embrionario donde la ausencia de CCM2 y CCM3 provoca defectos en el saco vitelino[178,188,219], se encontraron afectados genes relacionados con el desarrollo de la placenta (**Figura R8B**). Además, los genes desregulados se encontraban relacionados con el proceso de diferenciación celular, lo que puede indicar alteraciones en la salida del estado pluripotente y especificación celular en ausencia de estos genes *Ccm* (**Figura R8B**).

Para entender mejor como afecta la ausencia de CCM2 y CCM3 a la desregulación de estos 121 genes compartidos, analicé su expresión relativa a ESCs WT. Este análisis comparativo de expresión diferencial reveló la existencia de dos grupos principales: (1) grupo con 58 genes que muestran la misma tendencia en la desregulación de su expresión entre las ESCs mutantes para *Ccm2* y *Ccm3* y abarcan dos clases (1 y 2); y (2) grupo con 63 genes que muestran tendencias opuestas en su expresión entre las células mutantes y abarcan dos clases adicionales (3 y 4) (**Figura R8C**). La clase 1 contiene aquellos genes cuya expresión se encuentra disminuida en ausencia de *Ccm2* y *Ccm3* (p. ej. *Eomes* y *Ptk2b*), mientras que los de la clase 2 son reprimidos por estas dos proteínas (p. ej. *Ddah1* y *Plk2*) (**Figuras R8C** y **R8D**). La clase 3 (p. ej. *Tbx3* y *Drmt1*) y clase 4 (p. ej. *Usp25* y *Efnb2*) engloban genes cuya expresión se ve incrementada solo en las ESCs mutantes para *Ccm2* o *Ccm3*, respectivamente (**Figuras R8C** y **R8D**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

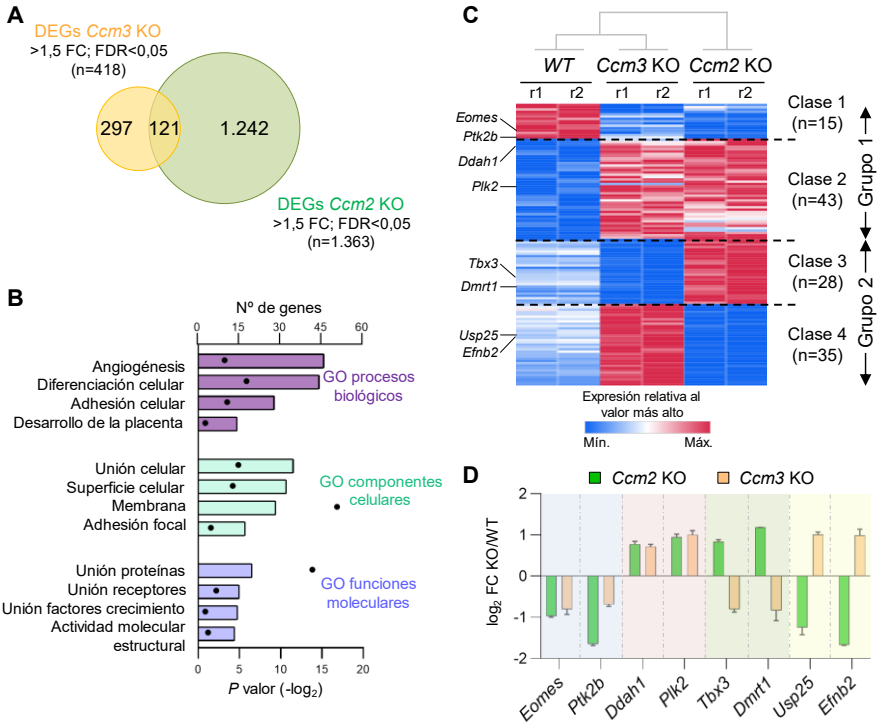


Figura R8. Análisis comparativo del perfil transcripcional alterado en ausencia de CCM2 y CCM3. A) Representación mediante diagrama de Venn de los genes diferencialmente expresados (DEGs) en las ESCs KO para *Ccm2* y *Ccm3* con respecto a las ESCs WT. **B)** Análisis de ontología génica (GO) empleando los 121 DEGs comunes en ESCs KO para *Ccm2* y *Ccm3*. Se muestra el enriquecimiento de los principales procesos afectados de manera significativa en las categorías de "procesos biológicos", "componentes celulares" y "funciones moleculares" a través de la significancia y el número de genes desregulados involucrados en cada término. Las barras representan el *P* valor y los puntos el número de genes. **C)** Representación mediante mapa de calor de la expresión génica de los replicados de ESCs WT y KO para *Ccm2* y *Ccm3* acoplado a una agrupación jerárquica. Se emplearon los datos de RNA-seq de los 121 DEGs comunes identificados en el panel **A**. Se representan los genes con la misma tendencia de desregulación (Grupo 1) y con tendencia opuesta (Grupo 2). Dentro de cada grupo, se ordenaron los genes en clases dependiendo de su represión o activación. **D)** Representación mediante gráfica de barras del valor del cambio de expresión (log_2 FC) de genes representativos de cada clase resaltados en el panel **C**. Los datos se presentan como media ± s.e.m.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

En su totalidad, estos resultados muestran como a pesar de que la ausencia de las CCM no compromete el mantenimiento de un estado pluripotente, sí lo altera a nivel transcripcional en comparación a las ESCs WT cultivadas en presencia de suero y LIF, lo que podría afectar al espectro de estados alternativos de pluripotencia que coexisten bajo estas condiciones de cultivo estándar. Por otro lado, mis resultados analizando la desregulación transcripcional en las ESCs KO para *Ccm2* y *Ccm3* revelan la alteración de procesos relacionados con la salida del estado celular pluripotente durante la diferenciación celular y procesos relacionados a la patología de las malformaciones cavernosas cerebrales asociada a la pérdida de función CCM, entre los que se encuentran la función endotelial y la angiogénesis. Por último, cabe destacar que el 90% de los genes desregulados por la ausencia de CCM2 son únicos de esta condición, lo que sugiere la posible existencia de funciones independientes de CCM2 con respecto a las otras dos proteínas CCM en el contexto de pluripotencia.

5.1.6 La ausencia de CCM2 promueve pluripotencia *naïve*

Puesto que mis resultados sugieren que las ESCs sin CCM1, CCM2 o CCM3 podrían presentar afectado el espectro de estados alternativos pluripotentes que existen en condiciones de cultivo estándar (SL), lo siguiente que me propuse fue explorar con más detalle esta posibilidad. Primero comprobé la expresión de los tres genes en los principales estados celulares que cubren el espectro pluripotente: *naïve*, metaestable, formativo y *primed*. Se observó que solo *Ccm2* cambia su expresión de forma significativa, viéndose reducida en el estado *naïve* (**Figura R9A**). De acuerdo con la literatura, los niveles de expresión de *Oct4* no variaron entre los estados alternativos pluripotentes, *Nanog* se encontró regulado al alza en el estado pluripotente *naïve* y *Otx2* mostró sus niveles más elevados en el estado pluripotente formativo[17,45,49,50]. Dado que los distintos estados pluripotentes en cultivo representan un continuo del desarrollo embrionario, decidí analizar a qué punto del espectro *in vivo* se parecen las ESCs WT y KO para cada gen *Ccm*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cultivadas en condiciones estándar (SL). Para ello, tomé ventaja del estudio de Boroviack y colaboradores, en el que caracterizan mediante RNA-seq diferentes estados celulares durante el desarrollo embrionario temprano[275]. Así, se pudo observar mediante un análisis PCA que, al igual que las ESCs WT, las células KO para *Ccm1* y *Ccm3* se agrupaban con las células del epiblasto a día E4.5, mientras que las células *Ccm2* KO se agrupaban con las células de la masa celular interna (ICM) (**Figura R9B**). Como se ha descrito previamente, el estado pluripotente *naïve* en cultivo mimetiza al estado pluripotente de las células de la ICM[17], lo que sugiere que sin CCM2 las ESCs cultivadas en presencia de suero y LIF (SL) se encuentran en un estado celular más similar a la pluripotencia *naïve*.

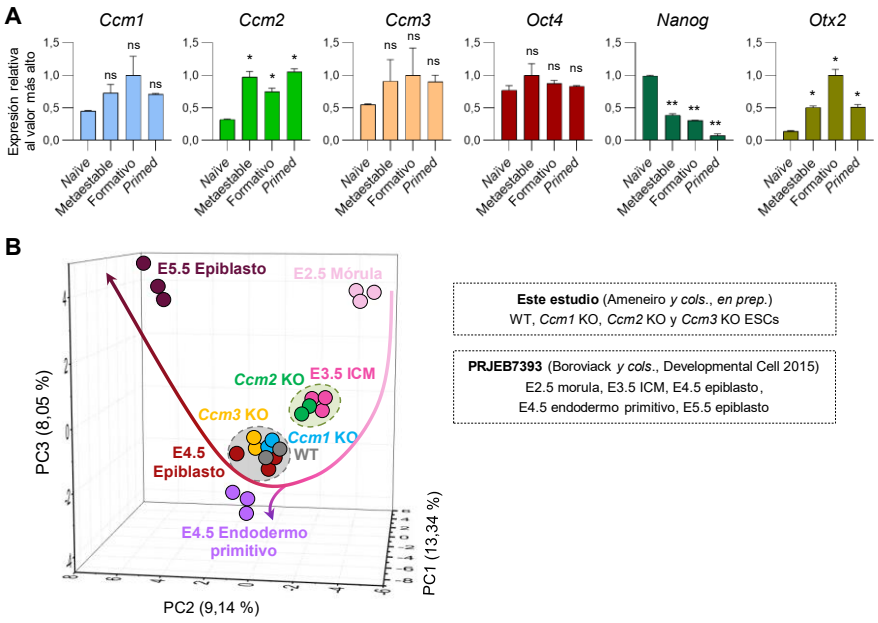


Figura R9. Las ESCs KO para *Ccm2* son transcripcionalmente similares a las células de la masa interna. A) Representación mediante gráficas de barras de la expresión relativa de los genes *Ccm* y de marcadores de pluripotencia (*Oct4*, *Nanog* y *Otx2*) en los estados de pluripotencia *naïve*, metaestable, formativo y *primed*. Los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre cada estado y el *naïve*. *p<0,05; **p<0,01; ns: no significativo. **B)** Análisis mediante PCA de los perfiles transcripcionales de las células

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ESCs WT y KO para *Ccm1*, *Ccm2* y *Ccm3*, y los estados tempranos de desarrollo embrionario E2.5 mórula, E3.5 ICM, E4.5 epiblasto, E4.5 endodermo primitivo y E5.5 epiblasto. Se empleó la lista de genes desregulados por la ausencia de CCM1, CCM2 o CCM3. Se utilizaron los datos de RNA-seq generados en esta tesis y en el estudio de Boroviack y colaboradores (PRJEB7393)[275]. El porcentaje de variación explicado por cada componente principal (PC) se indica en el eje correspondiente.

Este cambio en el estado pluripotente de las células KO para *Ccm2*, junto a mis datos anteriores donde se muestra la singularidad de esta línea celular en comparación con los otros KO y su contraparte WT (**Figuras R6-R9**), me llevó a centrar mis análisis en caracterizar en profundidad el papel de CCM2 en pluripotencia. Lo primero que hice fue un análisis de enriquecimiento de conjuntos de genes (GSEA, del inglés *gene set enrichment analysis*) para dos listas independientes definidas en dos estudios previos que caracterizaban la firma génica del estado pluripotente *naïve*[40,73]. Mis resultados revelaron que los niveles de expresión del transcriptoma de las células *Ccm2* KO se encuentran enriquecidos para las dos listas de genes con relación a las ESCs WT (**Figura R10A**). En ambos estudios se reprograman las ESCs desde un estado pluripotente metaestable o heterogéneo en SL mediante el uso de medio definido *naïve* caracterizado por la ausencia de suero y presencia de LIF junto con los inhibidores de las quinasas MEK y GSK3β (de aquí en adelante medio 2iL, para más detalle ver apartado 4.3). Al comparar las dos listas de genes, se identificaron 558 genes comunes a ambas, constituyendo una firma génica *naïve* más robusta (**Figura R10B**). En sintonía con los resultados anteriores, el análisis de la expresión de 32 genes *naïve* representativos de esta lista mostró un incremento significativo en las células KO para *Ccm2* con respecto a las ESCs WT o la ausencia de cada una de las otras dos proteínas CCM (**Figura R10C**). En conjunto, estos resultados muestran que la pérdida de CCM2 en ESCs cultivadas en condiciones estándar (SL) promueve un cambio transcripcional más similar al estado alternativo pluripotente *naïve*.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

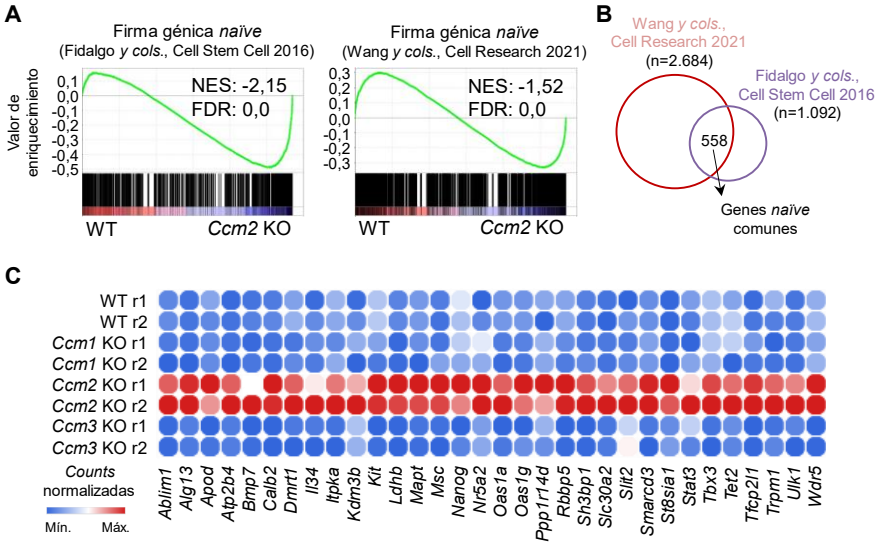


Figura R10. CCM2 reprime la expresión de genes asociados con la pluripotencia *naïve*. A) Análisis de enriquecimiento de conjunto de genes (GSEA) de los datos de RNA-seq de ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo contra firmas génicas *naïve* determinadas en dos estudios independientes (Fidalgo y colaboradores[73], izquierda; Wang y colaboradores[40], derecha). Se muestra la puntuación de enriquecimiento normalizada (NES, del inglés *normalized enrichment score*) y la tasa de descubrimientos falsos (FDR, del inglés *false discovery rate*). **B)** Representación mediante diagrama de Venn de las dos firmas génicas *naïve* mostrando los genes comunes. **C)** Representación mediante mapa de calor de la expresión génica de los replicados de ESCs WT y KO para *Ccm1*, *Ccm2* y *Ccm3*. Se emplearon los datos de *counts* normalizadas del RNA-seq de 32 genes representativos *naïve* comunes a las dos listas.

5.1.7 CCM2 contribuye a la salida de la pluripotencia *naïve*

Mis datos muestran que CCM2 interviene en la regulación del control de la expresión génica en pluripotencia. Además, los resultados obtenidos sugieren que esta proteína podría tener un papel relevante en la salida de la pluripotencia *naïve* y en la activación de programas de diferenciación. Con el fin de explorar estas posibilidades, en primer lugar me propuse determinar la homogeneidad en la expresión de NANOG en ausencia de CCM2, ya que las ESCs cultivadas en

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

presencia de suero y LIF (SL) representan un cultivo heterogéneo y muestran un patrón bimodal en la expresión de ciertos marcadores, entre los que se incluye este regulador esencial de la pluripotencia, cuyos niveles son más elevados en el estado celular pluripotente *naïve*[17,45]. Para ello, empleé una línea *knock-in* de ESCs que expresa la proteína fluorescente verde GFP bajo el control del promotor de *Nanog* (*Nanog*-GFP ESCs) proporcionada por el Dr. Konrad Hochedlinger (Harvard Univ., USA). Estas células se transdujeron usando vectores lentivirales conteniendo shARNs para silenciar la expresión de *Ccm2* o *Luciferasa* como control (**Figura R11A**). La eficiencia del silenciamiento de *Ccm2* se comprobó mediante qPCR (**Figura R11B**). Seguidamente, se evaluó el impacto de la pérdida de CCM2 en la distribución de NANOG en ESCs cultivadas en medio SL a partir de la detección de los niveles de GFP mediante FACS. Los resultados obtenidos mostraron que el silenciamiento de *Ccm2* incrementa notablemente el porcentaje de la actividad del reportero GFP (63,2%) en comparación con las ESCs transducidas con el control shLuci (48,7%) (**Figura R11C**).

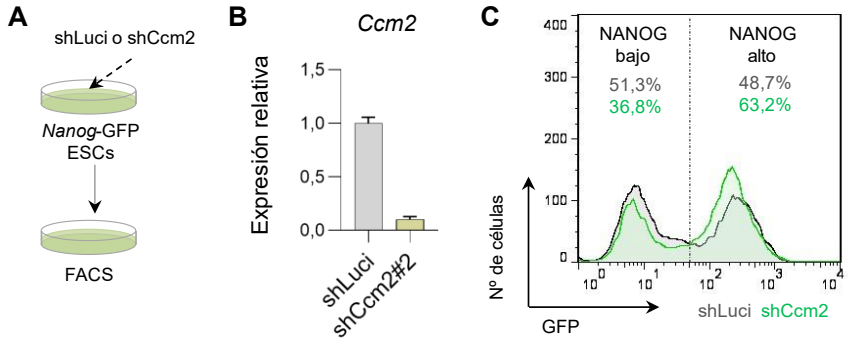


Figura R11. La eliminación de CCM2 conlleva un aumento de la población de ESCs con niveles elevados de NANOG. A) Representación esquemática de la estrategia seguida para analizar los niveles de NANOG. ESCs que expresan GFP bajo el control del promotor de *Nanog* se transdujeron con vectores lentivirales que contienen shARNs contra *Ccm2* o *Luciferasa* como control. La cuantificación de la expresión de NANOG se realizó mediante la medición de la intensidad de fluorescencia de GFP por citometría de flujo (FACS). **B)** Representación mediante gráfica de barras de la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

eficiencia de silenciamiento de *Ccm2* en ESCs *Nanog*-GFP transducidas con shARNs contra *Ccm2* o *Luciferasa* determinado por qPCR. n=3 experimentos independientes. Los datos se presentan como media ± s.e.m. **C)** Análisis mediante FACS de los niveles de NANOG en ESCs *Nanog*-GFP transducidas con shARNs contra *Ccm2* o *Luciferasa* en condiciones de cultivo estándar (SL). Se muestra el porcentaje de células con niveles de NANOG bajos y altos para las dos condiciones.

Seguidamente, decidí analizar si la pérdida de CCM2 favorece la retención de rasgos morfológicos característicos de la pluripotencia *naïve*. Con este fin, estimulé la diferenciación de las ESCs WT y ESCs KO durante 5 días retirando la citoquina LIF en el medio de cultivo estándar de pluripotencia. A continuación, forcé la reprogramación de ambos tipos celulares hacia la pluripotencia *naïve* mediante la realización de un ensayo de formación de colonias en medio de cultivo *naïve* (2iL) (**Figura R12A**). La decisión celular de comenzar un proceso de diferenciación no es reversible sin una estimulación concreta, por lo que toda célula que haya emprendido el cambio de identidad celular tendrá comprometida su capacidad de ser desdiferenciada hacia el estado celular pluripotente. De manera destacada, se observó que tras 5 días en medio de reprogramación 2iL, las células *Ccm2* KO fueron capaces de dar lugar a un mayor número de colonias no diferenciadas, con la característica morfología compacta del estado pluripotente *naïve*, en comparación a las células WT (**Figuras R12B** y **R12C**). Estos datos sugieren que la ausencia de CCM2 dificulta o retrasa el cambio de identidad celular durante la salida de pluripotencia.

Estos resultados, junto con el perfil transcripcional celular identificado en la sección anterior, apoyan que CCM2 desempeña un papel importante en la transición entre estados alternativos de pluripotencia.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

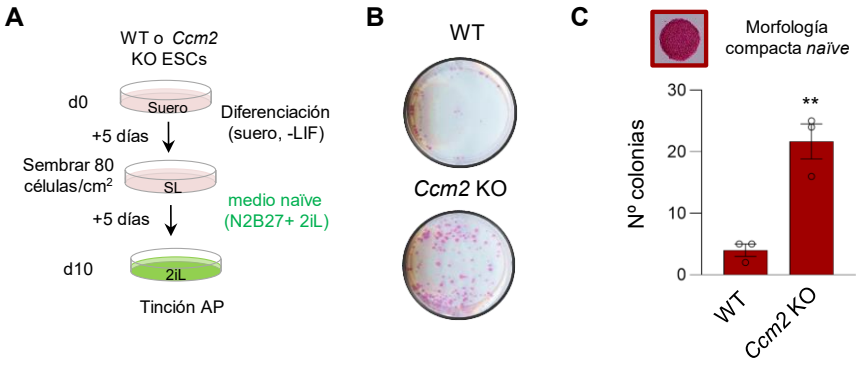


Figura R12. La ausencia de CCM2 favorece la reprogramación a pluripotencia *naïve* tras diferenciación celular. A) Representación esquemática de la estrategia seguida para realizar un ensayo de formación de colonias a partir de células diferenciadas. Se estimuló la diferenciación de ESCs WT o KO para *Ccm2* mediante su cultivo en ausencia de LIF durante 5 días. A continuación, se sembraron 80 células/cm^2 en condiciones estándar de cultivo (SL). Al día siguiente se cambió el medio por medio *naïve* (N2B27+2iL). Tras otros 5 días se realizó una tinción AP y se contaron las colonias formadas con morfología compacta característica del estado pluripotente *naïve*. **B)** Imágenes representativas de la tinción AP para las dos condiciones. **C)** (Arriba) Imagen representativa de una colonia con morfología compacta *naïve*. (Abajo) Representación mediante gráfica de barras del número total de colonias con morfología *naïve* para las dos condiciones. n=3 experimentos independientes. Los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico. **p<0,01.

5.1.8 CCM2 es necesario para establecer y mantener la pluripotencia *primed*

El posible defecto o retraso que la eliminación de CCM2 provoca en la salida de pluripotencia (**Figura R12)**, me llevó a analizar el efecto que la depleción de esta proteína puede ejercer durante el cambio de identidad celular entre distintos estados alternativos pluripotentes. Para ello, tomé ventaja de que las ESCs de origen murino cultivadas en presencia de suero y LIF (SL), que representan un estado heterogéneo o metaestable, pueden ser reprogramadas en dos estados pluripotentes interconvertibles y definidos mediante la simple modulación de distintas vías de señalización. En concreto, tal como se

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ha comentado anteriormente, el cultivo en medio sin suero (N2B27) y con presencia de los inhibidores de MEK y GSK3β y la citoquina LIF (2iL) promueve un estado pluripotente *naïve* que imita al de la masa celular interna (ICM) del blastocisto[17,29]. Por otro lado, se puede inducir un estado más avanzado y comprometido de pluripotencia o *primed* que se asemeja al que se encuentra en las células del epiblasto tras la implantación uterina mediante medio sin suero (N2B27) que contenga las citoquinas de factor de crecimiento de fibroblastos 2 (FGF2) y ACTIVINA A (FA)[17,38]. Teniendo esto en cuenta, analicé el papel de CCM2 en la reprogramación de las ESCs cultivadas en SL hacia los estados pluripotentes definidos *naïve* y *primed*. Con el fin de poder cuantificar la eficiencia de la reprogramación de ESCs WT o KO hacia los distintos estados celulares pluripotentes, procedí a sembrar cada tipo celular en densidad clonal sobre placas de cultivo recubiertas con el apropiado sustrato (gelatina y fibronectina para *naïve* y *primed*, respectivamente) y medio de cultivo definido (**Figura R13A**). Los resultados obtenidos a lo largo del presente trabajo sugieren que las células KO para *Ccm2* adquieren una identidad similar a la *naïve*, por lo que cabría esperar que la ausencia de esta proteína facilite la reprogramación en medio *naïve* 2iL, y dificulte o retrase el cambio de identidad celular en medio *primed* FA. Cabe resaltar que la pluripotencia *naïve* se caracterizan por una morfología compacta y regular de las colonias pluripotentes, mientras que las colonias pluripotentes *primed* muestran una morfología plana e irregular[17]. Mis resultados mostraron que, de acuerdo con mi hipótesis, la ausencia de CCM2 fue suficiente para incrementar la eficiencia del proceso de reprogramación al estado pluripotente *naïve*, determinado por el aumento del número de colonias características de este estado, mientras que dificultó el cambio de identidad hacia la pluripotencia *primed*, reflejado por la retención de colonias compactas (**Figuras R13B** y **R13C**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

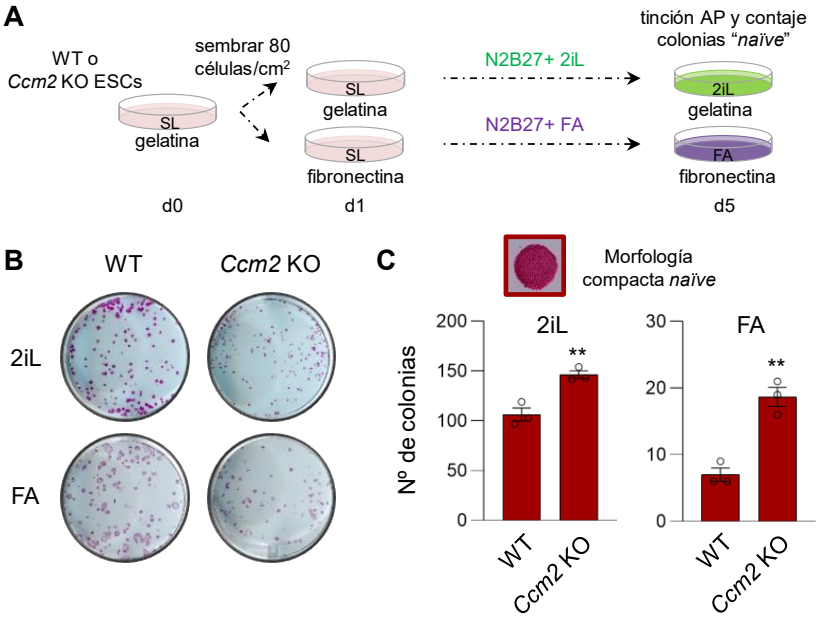


Figura R13. La falta de CCM2 dificulta la obtención del estado pluripotente *primed*. A) Representación esquemática de la estrategia seguida para analizar la eficiencia de reprogramación de células metaestables (en condiciones estándar de cultivo) hacia células *naïve* y *primed*. Se sembraron 80 células/cm^2 en condiciones estándar de cultivo (SL) en placas tratadas con gelatina o fibronectina. Al día siguiente se cambió el medio por medio *naïve* (N2B27+2iL, en gelatina) o medio *primed* (N2B27+FA, en fibronectina). Tras 4 días se realizó una tinción AP y se contaron las colonias formadas con morfología compacta *naïve*. **B)** Imágenes representativas de la tinción AP para todas las condiciones. **C)** (Arriba) Imagen representativa de una colonia con morfología compacta *naïve*. (Abajo) Representación mediante gráfica de barras del número total de colonias con morfología *naïve* para todas las condiciones. n=3 experimentos independientes. Los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico. **p<0,01.

Con el fin de descartar que los resultados observados se deban a defectos en la proliferación de las células cultivadas en los medios 2iL o FA, decidí realizar un ensayo de competición de la proliferación celular. Para ello, generé co-cultivos de ESCs WT o KO para *Ccm2* con una línea reportera de ESCs WT, que expresan de manera constitutiva el marcador fluorescente tdTomato, en una proporción 80-20%

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

respectivamente, y los mantuve durante 4 días en medio *naïve* o *primed*. A días 1 y 4 cuantifiqué el porcentaje de células tdTomato positivas mediante FACS (**Figura R14A**). Mis datos mostraron que ninguna de las condiciones celulares analizadas presentaba defectos de proliferación dependientes del fondo genético de las ESCs analizadas en cada condición de cultivo en medio definido, al no variar de gran manera el porcentaje inicial 80-20% (**Figura R14B**). Así, podemos confirmar que el aumento o detrimento observados en la eficiencia de reprogramación de las ESCs KO para *Ccm2* a *naïve* y *primed*, respectivamente, no son consecuencia de una modificación de los patrones de proliferación debido a las condiciones específicas del cultivo empleado durante la reprogramación.

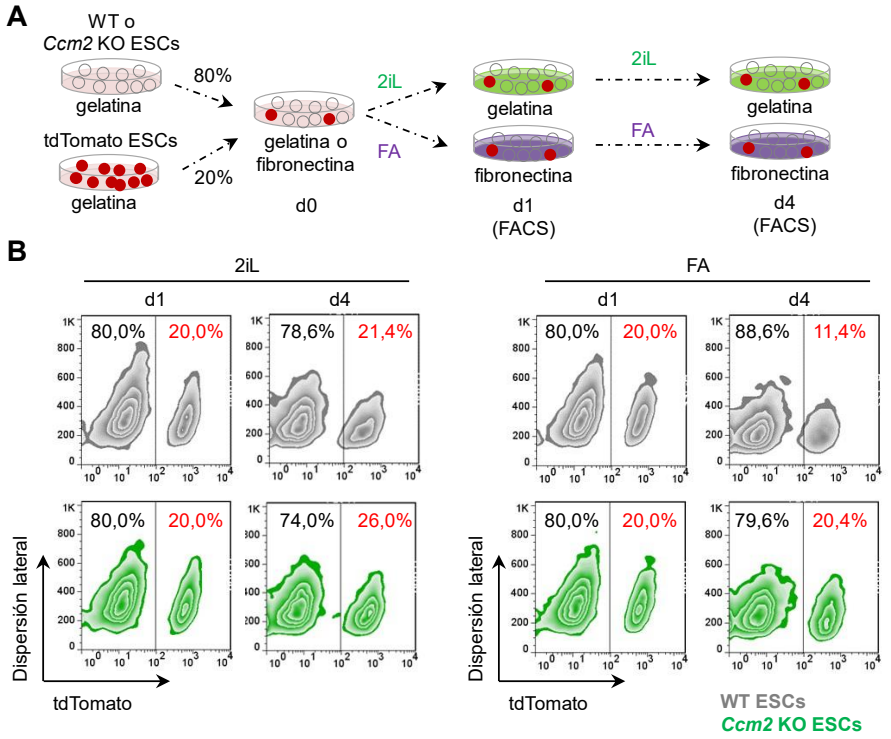


Figura R14. La pérdida de función de CCM2 no altera la proliferación celular durante la reprogramación a estados alternativos pluripotentes. A)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Representación esquemática de la estrategia seguida para analizar la proliferación celular en medio *naïve* o *primed*. Se sembraron ESCs WT o KO para *Ccm2* mezcladas con ESCs que expresan constitutivamente el marcador fluorescente tdTomato con una proporción 80-20%. Se emplearon condiciones estándar de cultivo (SL) y placas tratadas con gelatina o fibronectina. Al día siguiente se cambió el medio por medio *naïve* (N2B27+2iL, en gelatina) o medio *primed* (N2B27+FA, en fibronectina). A día 1 y 4 se cuantificó el porcentaje de cada línea celular mediante citometría de flujo. **B)** Representación de la cuantificación del porcentaje de células positivas y negativas para la expresión de tdTomato mediante FACS para todas las condiciones a días 1 y 4.

Estos resultados me llevaron preguntarme que ocurriría si eliminásemos CCM2 en células pluripotentes que ya se encuentran en el estado celular *primed*. Con este fin, transduje células troncales derivadas del epiblasto tardío (EpiSCs), que representan el estado pluripotente *primed*[17], con shARNs contra *Ccm2* o *Luciferasa* como control. Tras mantener dos líneas independientes de EpiSCs (EpiSC9 y OEC2) en medio *primed* FA durante 5 o 9 días respectivamente, analicé el efecto de la pérdida de CCM2 (**Figura R15A**). De manera destacable, se observó que la ausencia de CCM2 compromete fenotípicamente la morfología plana e irregular que resulta característica en las colonias pluripotentes en estado *primed*, dando como resultado la aparición de colonias con apariencia compacta y redondeada típica de la pluripotencia *naïve* (**Figura R15B).**

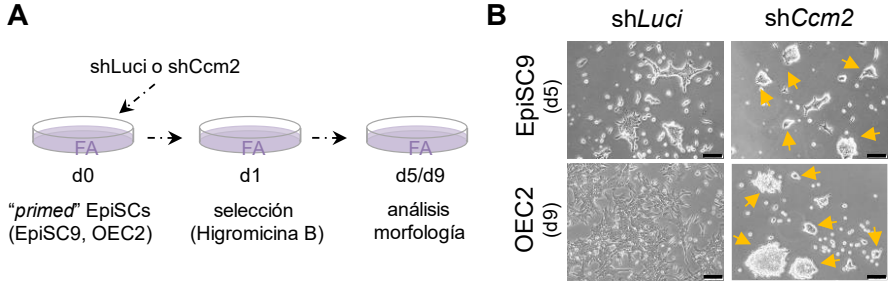


Figura R15. CCM2 es necesario para el mantenimiento del estado pluripotente *primed*. A) Representación esquemática de la estrategia seguida para analizar el efecto de la pérdida de *Ccm2* en EpiSCs. Dos líneas independientes de células pluripotentes en estado *primed* (EpiSC9 y OEC2) cultivadas en medio FA se transdujeron con vectores lentivirales que contienen shARNs contra *Ccm2* o

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Luciferasa como control y un gen de resistencia a higromicina. Se seleccionaron las células con antibiótico y a los días 5 y 9 se analizó la morfología de las colonias mediante microscopía. **B)** Imágenes representativas en campo claro de las condiciones indicadas. Las flechas señalan el cambio morfológico de las colonias en las EpiSCs transducidas con shCcm2. Las barras representan 100 µm.

En conjunto, mis resultados muestran que la ausencia de CCM2 altera el estado pluripotente metaestable de las ESCs en condiciones estándar, acercándolas al estado pluripotente *naïve* tanto a nivel transcripcional como fenotípico. Además, dificulta el mantenimiento del estado pluripotente *primed*, más restringido y preparado para activar los procesos de diferenciación. Por otro lado, los datos sugieren que sin CCM2, las células podrían presentar defectos durante la salida del estado pluripotente, lo que será analizado en la sección 5.3.

5.2 ELUCIDAR LOS PROGRAMAS MOLECULARES GOBERNADOS POR CCM2 EN PLURIPOTENCIA

5.2.1 CCM2 no interviene en el control de la expresión génica a nivel de cromatina

Tras comprobar que sin CCM2 las células entran en un estado pluripotente similar al *naïve*, la siguiente pregunta que me propuse contestar fue cómo actúa CCM2 en esta transición. Para ello, lo primero que hice fue analizar la localización subcelular de esta proteína. Dado que no disponemos de un anticuerpo optimizado para la detección de CCM2 en células murinas, decidí generar una línea de rescate (^{3xFL}CCM2) en las ESCs KO para *Ccm2* mediante el uso del sistema de transposones PiggyBac que codifica para una versión WT de CCM2 marcada con tres epítopos Flag en su extremo amino-terminal (**Figura R16A**). La validación de esta línea de rescate para CCM2 en ESCs KO se realizó mediante western blot utilizando un anticuerpo específico contra FLAG (**Figura R16B**). Aunque en células endoteliales se ha descrito que CCM2 se encuentra únicamente en el citoplasma[171], mis resultados muestran que en células pluripotentes se encuentra tanto en el citoplasma como en el núcleo celular (**Figura R16C**). Llevando a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cabo un análisis más detallado en el que se distinguen las fracciones de citoplasma, soluble nuclear y unida a cromatina, se observa que CCM2 está presente en las tres localizaciones subcelulares, aunque preferentemente en la citoplasmática (**Figura R16D**).

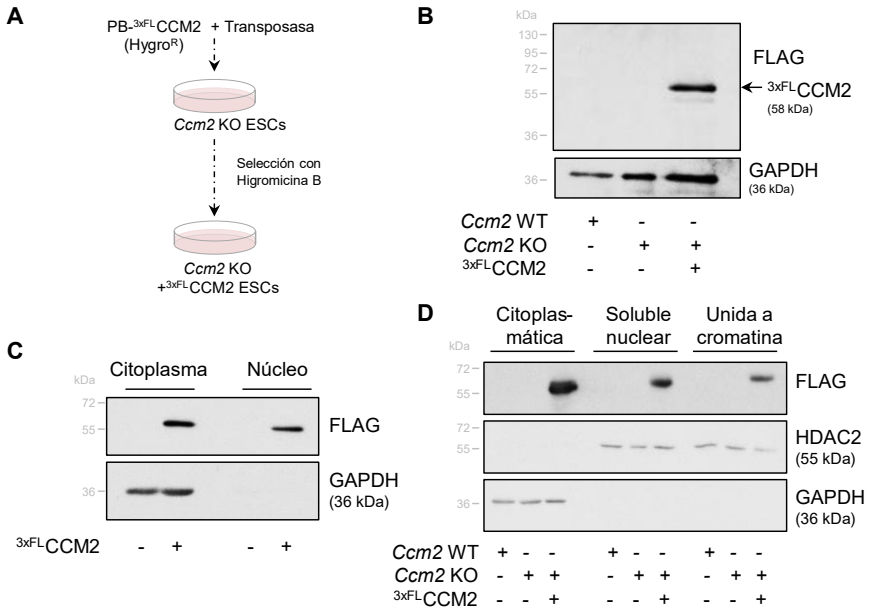


Figura R16. CCM2 se encuentra presente en la cromatina de ESCs. A) Representación esquemática de la estrategia seguida para generar una línea de rescate de CCM2 en ESCs KO para este gen. Se transfectaron ESCs KO para *Ccm2* con el sistema de sobreexpresión basado en transposones PiggyBac conteniendo el CDS de *Ccm2* unido a tres epítopos Flag y un gen de resistencia a higromicina. Se seleccionaron las células con antibiótico para establecer una línea de ESCs homogénea expresando ^{3xFL}CCM2. **B)** Validación de la sobreexpresión de ^{3xFL}CCM2 mediante western blot empleando un anticuerpo contra FLAG. GAPDH se empleó como control de carga. **C)** Western blot para analizar la localización de ^{3xFL}CCM2 en las fracciones citoplasmática y nuclear de ESCs WT y KO para *Ccm2* rescatadas con ^{3xFL}CCM2. GAPDH se empleó como control de carga y como control de fracción citoplasmática. **D)** Western blot para analizar la localización de ^{3xFL}CCM2 en las fracciones citoplasmática, nuclear soluble y nuclear unida a cromatina de ESCs WT, *Ccm2* KO y *Ccm2* KO + ^{3xFL}CCM2. GAPDH se empleó como control de carga y como control de fracción citoplasmática. HDAC2 se empleó como control de fracciones nucleares.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Una vez comprobado que CCM2 está presente en la cromatina de las ESCs, analicé la posibilidad de que esta proteína intervenga de manera directa en el control de la expresión de los genes desregulados en su ausencia (**Figura R7A**). Para caracterizar potenciales dianas transcripcionales de CCM2 en células pluripotentes, examiné los lugares de unión de CCM2 en el genoma a nivel global mediante inmunoprecipitación de la cromatina seguida de secuenciación (ChIP-seq, por sus siglas en inglés) (**Figura R17A**). Antes de proceder con la secuenciación masiva se verificó la inmunoprecipitación de ^{3xFL}CCM2 (**Figura R17B**). El análisis *in silico* de los resultados obtenidos identificó solamente 209 picos de unión de CCM2 en el genoma, cercanos a 157 genes (**Figura R17C**). Seguidamente, analicé la unión de CCM2 en los 1.363 genes desregulados en las ESCs KO para *Ccm2*, tanto activados como reprimidos (FC>1,5 y FDR<0,05), en las proximidades del codón de inicio de la transcripción (TSS; ± 3 kb) mediante gráficos de densidad y mapas de calor. Mientras que el análisis de ChIP-seqs publicados en ESCs para OCT4, SOX2 y NANOG[276] mostró un claro enriquecimiento de la presencia de estos reguladores esenciales de pluripotencia en los genes desregulados en ausencia de CCM2, la presencia de CCM2 sólo se observó en 6 genes activados (*Sh3bgrl*, *Ldhb-ps*, *Apba2*, *Btbd11*, *Ankrd44* y *Macroh2a1*) y en 1 gen reprimido (*Gm39469*) (**Figura R17D**). Estos resultados sugieren que CCM2 no participa de manera directa a nivel de cromatina en el control de la mayoría de los genes que se encuentran desregulados tras su pérdida de función en ESCs, ya que sólo el 0,5% de estos son dianas directas potenciales de esta proteína (**Figuras R17E** y **R17F**).

En conjunto, mis datos demuestran que, pese a su presencia en cromatina, CCM2 no actúa de manera directa interviniendo en la activación o represión transcripcional. Estos resultados sugieren que los cambios en la expresión de genes en las ESCs KO para *Ccm2*, que resultan en un transcriptoma más similar al que se encuentra en la pluripotencia *naïve*, son una consecuencia indirecta de la pérdida de función de esta proteína a nivel de cromatina.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

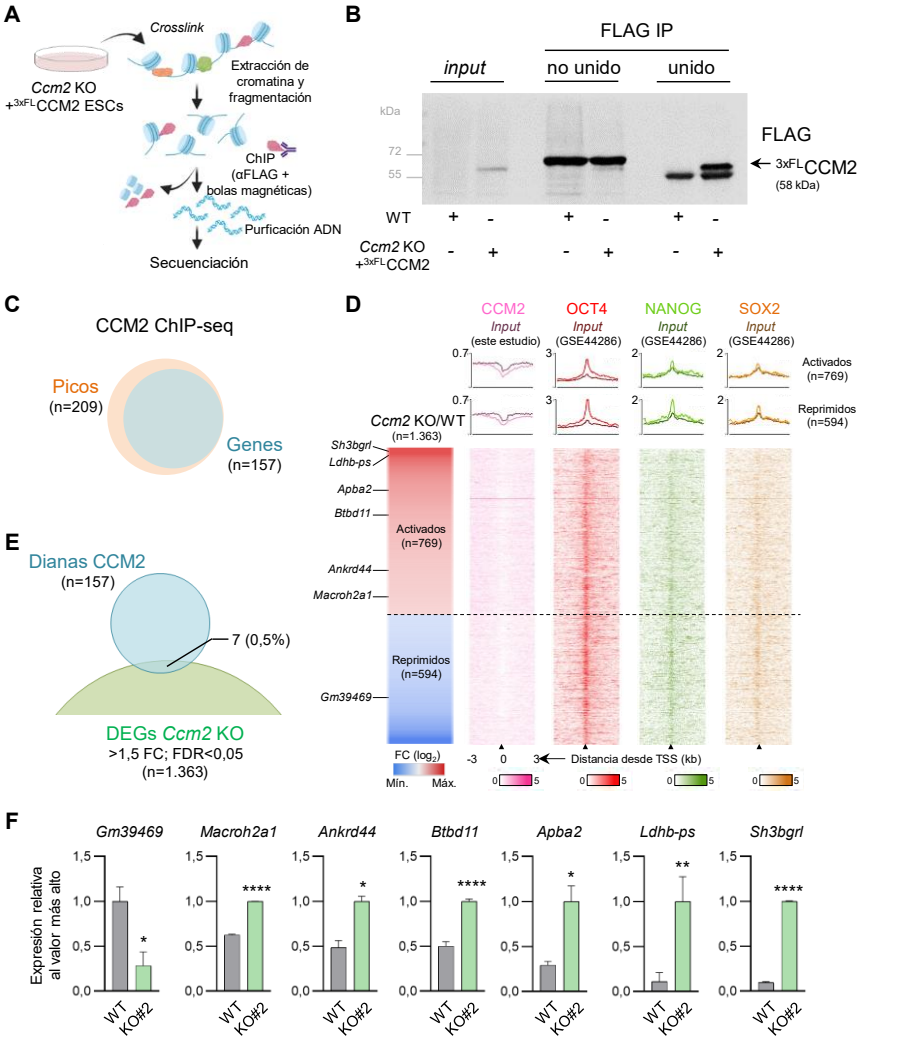


Figura R17. CCM2 no controla directamente la transcripción de los genes desregulados como resultado de su ausencia en ESCs. A) Representación esquemática de la estrategia de ChIP-seq seguida para analizar las dianas de CCM2 a nivel global en el genoma. Se extrajo la cromatina de las ESCs rescatadas para la ausencia de CCM2 (*Ccm2* KO +^{3xFL}CCM2) en condiciones estándar de cultivo tras la realización del proceso de entrecruzamiento (*crosslink*) para fijar las uniones existentes entre proteínas y ADN. Se fragmentó la cromatina mediante sonicación y se inmunoprecipitó (IP) ^{3xFL}CCM2 empleando un anticuerpo contra FLAG y bolas magnéticas para poder aislar la muestra precipitada. Se purificó el ADN recuperado

y se secuenció el material genómico inmunoprecipitado junto con material genómico sin inmunoprecipitar, que se utilizó como *input* para los análisis bioinformáticos posteriores. **B)** Validación de la IP mediante western blot empleando un anticuerpo contra FLAG en ESCs WT y *Ccm2* KO $+^{3xFL}$CCM2. **C)** Representación mediante diagrama de Venn del número de picos y genes identificados como dianas de CCM2. **D)** Representación mediante histograma (arriba) y mapa de calor (abajo) de la densidad de unión de CCM2, OCT4, NANOG y SOX2 a las regiones promotoras (± 3 kb del TSS) de los genes desregulados en ausencia de CCM2. Se emplearon los datos de ChIP-seq generados en esta tesis y los procedentes de ChIP-seqs previamente publicados por Whyte y colaboradores (GSE44286)[276] para OCT4, SOX2 y NANOG. Se muestran histogramas independientes para genes activados y reprimidos. Los genes desregulados se muestran ordenados de mayor a menor valor de cambio (FC) entre *Ccm2* KO y WT ESCs. **E)** Representación mediante diagrama de Venn de las dianas a nivel genómico identificadas de CCM2 y los genes desregulados en su ausencia. Se muestran 7 genes desregulados que son dianas, resaltados en el panel **D**. **F)** Representación mediante gráfica de barras de la expresión relativa de los 7 genes desregulados dianas de CCM2 en ESCs WT y KO para *Ccm2* calculada por qPCR. Los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre KO y WT. *p<0,05; **p<0,01; ****p<0,0001.

5.2.2 CCM2 es necesario para la activación de ERK1/2

Descartada la posibilidad de que CCM2 influya directamente a nivel de cromatina en el control de la expresión de los genes desregulados tras su pérdida, me propuse identificar los programas moleculares que resultan alterados en su ausencia en ESCs, y que como resultado alteran el transcriptoma de las células KO para *Ccm2*. Teniendo en cuenta que de los 1.363 transcritos desregulados en ausencia de CCM2 el 94% corresponde a genes codificantes de proteínas (**Figura R18A**), en primer lugar, procedí a realizar un análisis *in silico* de ontología génica de procesos biológicos a partir de estos genes diferencialmente expresados, identificando 162 procesos afectados significativamente (**Figura R18B**). De manera coherente con mis resultados anteriores y con las funciones previas descritas para CCM2 (sección 2.4), se identificaron entre los procesos más afectados significativamente aquellos relacionados con los contextos de diferenciación celular, desarrollo embrionario y biología endotelial (p. ej. angiogénesis, adhesión celular, desarrollo del sistema faríngeo, desarrollo embrionario en el útero y desarrollo de la placenta) (**Figura**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

R18B). Interesantemente, se encontraban alterados procesos biológicos relacionados con la fosforilación y la cascada de las quinasas de proteínas activadas por mitógenos (MAPK, del inglés *mitogen-activated protein kinases*) en ausencia de CCM2 (**Figura R18B**).

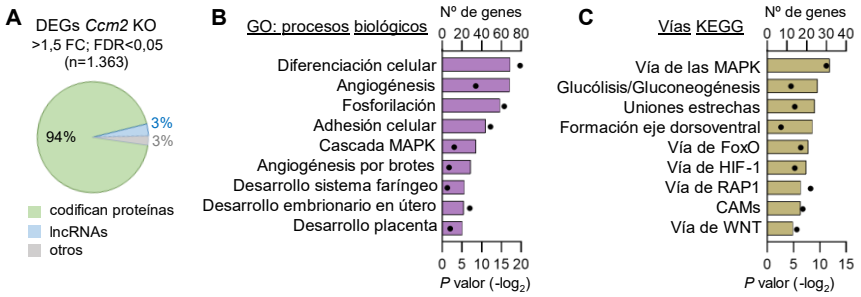


Figura R18. Identificación de procesos biológicos alterados y vías moleculares afectadas en *Ccm2* KO ESCs. A) Representación gráfica de genes diferencialmente expresados en ausencia de CCM2 indicando el porcentaje que constituye ARNs codificantes de proteínas, ARNs largos no codificantes y otros ARNs. **B)** Análisis de ontología génica (GO) de procesos biológicos mediante la herramienta DAVID empleando los genes desregulados codificantes de proteínas tras la eliminación de *Ccm2*. **C)** Análisis de las vías de señalización KEGG a través de DAVID de los genes desregulados en ausencia de CCM2 como en **B**. En **B** y **C** los términos representados se seleccionaron entre los más enriquecidos a nivel significativo. Las barras representan el *P* valor y los puntos el número de genes en cada categoría.

Luego, para obtener una visión más detallada a nivel molecular, realicé otro análisis *in silico* para identificar las vías KEGG (del inglés *kyoto encyclopedia of genes and genomes*) que se encuentran afectadas significativamente en las ESCs *Ccm2* KO (**Figura R18C**). De manera destacable, se observó que la vía de las MAPK es la que se encuentra más afectada significativamente, seguida por otras vías moleculares relacionadas con la correcta función endotelial, entre las que se encuentran la vía de FoxO, la vía de HIF-1 y la vía de RAP1 (**Figura R18C**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

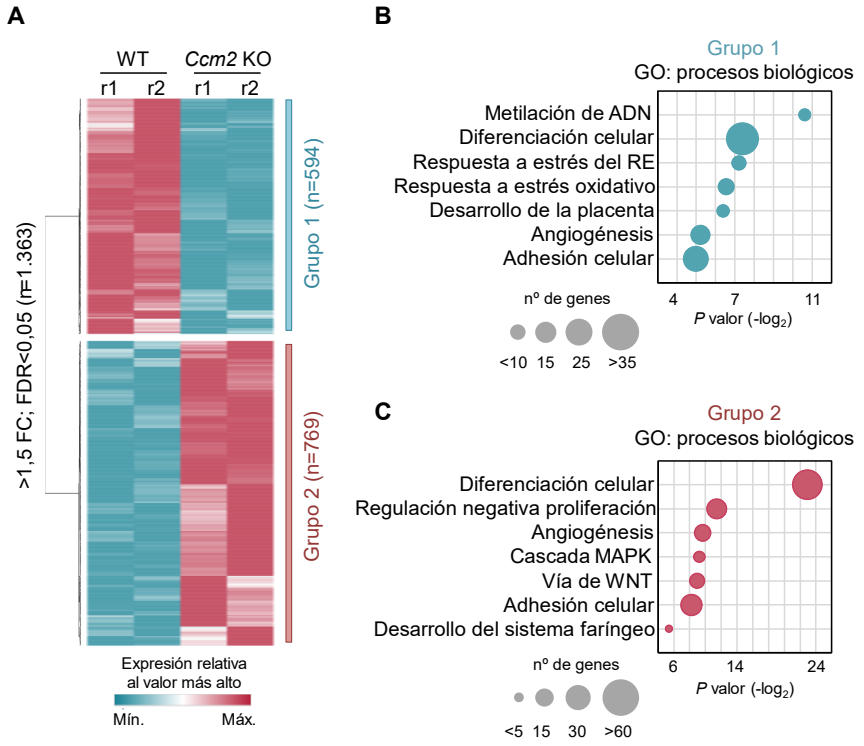


Figura R19. Identificación de procesos biológicos asociados a genes reprimidos y genes activados en ausencia de CCM2. A) Representación gráfica mediante mapa de calor de los genes diferencialmente expresados en ESCs KO para *Ccm2* con relación a ESCs WT. Se emplearon los datos de RNA-seq generados en este estudio, mostrando 594 genes reprimidos y 769 activados. **B)** Análisis de ontología génica (GO) de procesos biológicos empleando los 594 genes reprimidos tras la eliminación de *Ccm2*. Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de genes desregulados involucrados en cada proceso. **C)** Análisis de ontología génica (GO) de procesos biológicos empleando los 769 genes activados tras la eliminación de *Ccm2* igual que en **B**. En **B** y **C** los términos representados se seleccionaron entre los más enriquecidos a nivel significativo.

Lo siguiente que me propuse fue realizar un análisis similar al descrito anteriormente de ontología génica, pero esta vez centrándome en los genes reprimidos y en los genes activados en ausencia de CCM2. De los 1.363 genes diferencialmente expresados (FDR<0,05 y FC>1,5) en las ESCs KO para *Ccm2* comparado con las ESCs WT, el 43,6% de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

estos (n=594) se encontraron regulados a la baja, mientras que un 56,4% (n=769) mostraron un aumento en sus niveles de expresión (**Figura R19A**). Entre los procesos biológicos más afectados significativamente por los genes reprimidos en ausencia de CCM2, se encontraron procesos relacionados con la metilación de ADN, diferenciación celular, respuestas a estrés, y angiogénesis entre otros (**Figura R19B**). Por otro lado, con respecto a los relacionados con los genes activados en las ESCs *Ccm2* KO, se encontraron también procesos relacionados con la diferenciación celular y angiogénesis, así como los de las vías de señalización WNT y MAPK (**Figura R19C**).

La familia de las MAPKs consiste en 14 miembros que constituyen componentes clave para la transducción de señales que regulan procesos celulares, entre los que se encuentran la proliferación, la diferenciación o la muerte celular[277]. Esta familia de quinasas se clasifica en dos grupos, las MAPKs convencionales y las atípicas. En concreto, se ha mostrado que las MAPKs convencionales, que se clasifican a su vez en cuatro subfamilias [quinasas reguladas por señales extracelulares 1/2 (ERK1/2), quinasas N-terminal c-Jun (JNK1, 2 y 3), p38 (α, β, γ, δ) y quinasa 5 regulada por señales extracelulares (ERK5)], ejercen funciones relevantes en pluripotencia y en el control de los estados alternativos pluripotentes[278-281]. En base a esto, decidí analizar si la alteración de la cascada de las MAPK podría estar involucrada en el fenotipo similar a pluripotencia *naïve* observado en las ESCs *Ccm2* KO. En primer lugar, evalué el efecto de la pérdida de CCM2 en ESCs sobre los niveles de expresión de 826 genes que se encuentran regulados por estas quinasas o que participan en la regulación de éstas descritos previamente en otros contextos celulares (GO:0000165).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

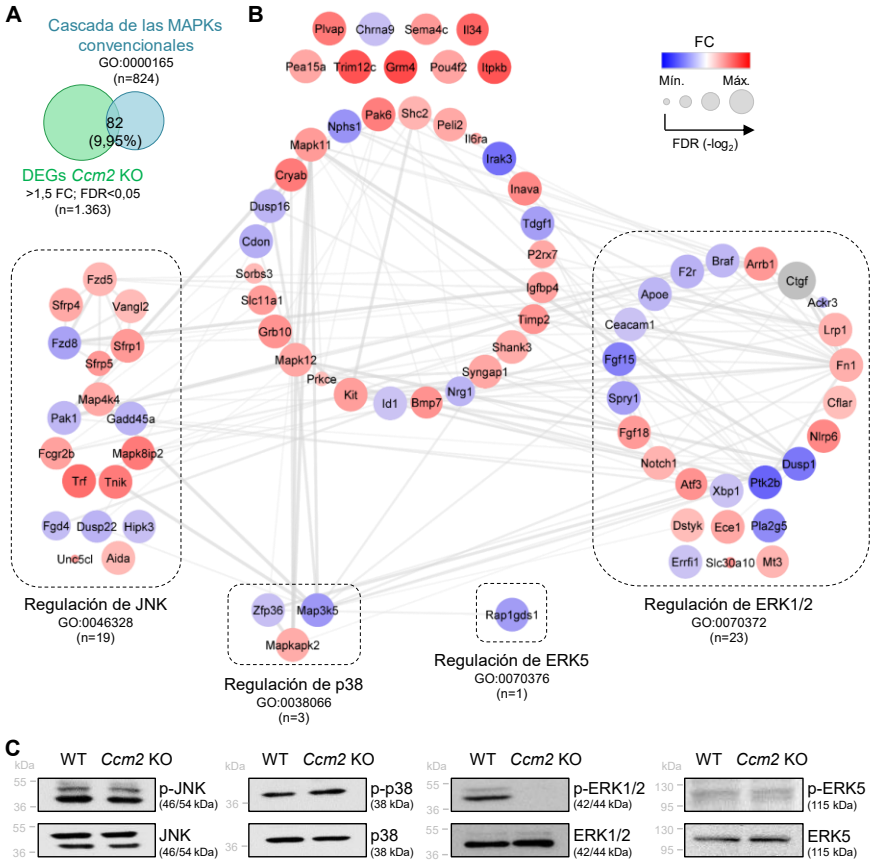


Figura R20. La ausencia de CCM2 compromete la activación de ERK1/2. A) Representación mediante diagrama de Venn de los genes involucrados en la cascada de MAPK (GO:0000165) y los genes desregulados en *Ccm2* KO ESCs (este estudio). **B)** Representación de las interconexiones validadas experimentalmente en diversos contextos celulares entre las 82 proteínas que codifican los genes desregulados empleando el software *Cytoscape*. Se resaltan aquellos genes relacionados con la regulación de las cuatro vías convencionales de la cascada MAPK (JNK, p38, ERK1/2 y ERK5). El valor de FC determina la intensidad de color de las burbujas (rojo equivale a activación; azul equivale a represión). El valor del -$\log_2$ FDR determina el tamaño de la burbuja (mayor tamaño equivale a mayor significancia). **C)** Western blot para detectar los niveles proteicos totales de JNK, p38, ERK1/2 y ERK5 y sus versiones fosforiladas que indican su estado de activación [p-JNK (Thr183/Tyr185), p-p38 (Tyr 182), p-ERK1/2 (Thr202/Tyr204 y Thr185/Tyr187) y p-ERK5 (Thr218/Tyr220)] en ESCs WT o KO para *Ccm2* en condiciones estándar de cultivo.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Notablemente, se observó que el 9,92% de estos genes se encuentran desregulados con un cambio superior al 50% y de manera significativa en ESCs *Ccm2* KO comparado con las ESCs WT (**Figura R20A**). Entre estos genes desregulados se identificaron 25 que intervienen en la regulación de la vía de ERK1/2, 18 en la vía de JNK, 3 en la de p38 y sólo uno en la vía de ERK5 (**Figura R20B**). Tras observar que estas cuatro cascadas presentan alteraciones en la expresión de reguladores específicos en ausencia de CCM2, procedí a analizar si CCM2 participa en la activación de estas MAPKs convencionales en ESCs. De manera destacada, los resultados indican que, mientras las vías de JKN, p38 y ERK5 no presentaron cambios apreciables en sus versiones fosforiladas activas, la pérdida de función de CCM2 conllevaba una reducción drástica de la activación de ERK1 y ERK2 que no es consecuencia de cambios en su abundancia a nivel de proteína (**Figura R20C**). Además, se observó que en las células pluripotentes en condiciones estándar de cultivo (SL) la isoforma más abundante correspondía a ERK2 en relación con ERK1 (banda 42 kDa y 44 kDa, respectivamente) (**Figura R20C**).

Estos datos indican que en el contexto de pluripotencia, la presencia de CCM2 es necesaria para la activación de ERK1/2, lo que contrasta con la literatura previa relacionada con el contexto endotelial y con la patología asociada a la pérdida de función de los genes *Ccm*, en donde se ha mostrado una relación directa entre la función de CCM2 y la modulación de la activación de la vía de ERK5, resultando la ausencia de CCM2 en un incremento de la activación[171,179,210,211].

5.2.3 CCM2 modula indirectamente la actividad de ERK1/2

Lo siguiente que me propuse fue investigar cómo la pérdida de función de CCM2 en ESCs compromete la activación de ERK1/2. A este respecto, la vía de señalización de ERK1/2 se encuentra altamente conservada desde microorganismos hasta mamíferos y su activación conlleva la recepción de señales desde distintos receptores de membrana hasta la activación de diversos factores de transcripción[282].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Por ejemplo, en ESCs la activación del receptor FGFR promueve un cambio conformacional activo de la proteína RAS de unión a GTP, lo que produce la activación de la quinasa RAF, que mediante fosforilación activa la quinasa MEK y ésta, a su vez, activa a las quinasas ERK1 y ERK2[32,143,144] (**Figura R21A**). Sin embargo, mis datos preliminares sugieren que la activación de RAF y MEF no se veía comprometida en ausencia de CCM2 en ESCs cultivadas en condiciones estándar (SL) (datos no mostrados). Por esto, decidí explorar si CCM2 está involucrada en la modulación de esta ruta a través de su interacción con componentes de la misma.

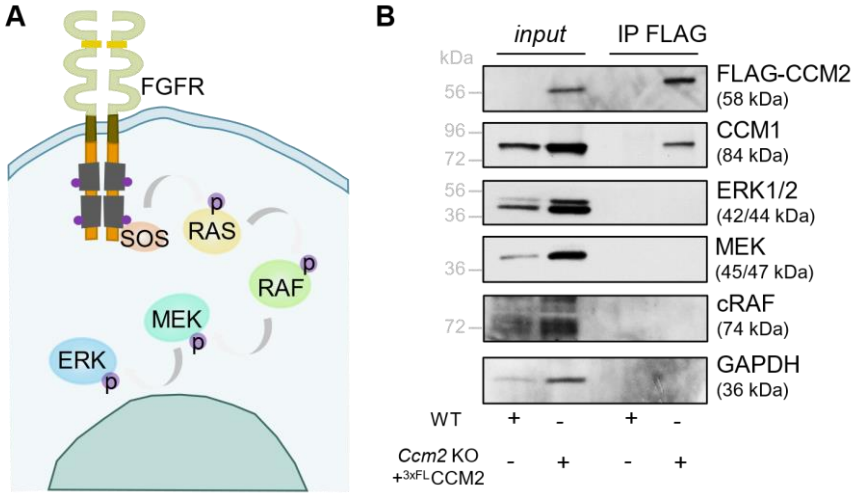


Figura R21. CCM2 no interacciona de manera directa con componentes de la vía ERK1/2. A) Representación esquemática simplificada de la vía de ERK1/2. Una vez activado el receptor de membrana FGFR se autofosforila y activa RAS mediante fosforilación a través de SOS. RAS a su vez fosforila y activa a RAF, ésta a MEK, y finalmente a ERK1/2 en el citoplasma. **B)** Análisis de la inmunoprecipitación (IP) de FLAG en ESCs WT y *Ccm2* KO $+^{3xFL}$CCM2. Detección mediante western blot de las proteínas CCM1, ERK1/2, MEK y cRAF. FLAG se empleó como validación de la IP y GAPDH como control de carga.

Con esta finalidad, realicé un ensayo de inmunoprecipitación empleando la línea de rescate generada anteriormente (*Ccm2* KO con sobreexpresión de ^{3xFL}CCM2). Esta aproximación se llevó a cabo a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

partir de muestras sin fijar, con el fin de detectar proteínas unidas de forma directa a CCM2, ya que las uniones secundarias o débiles se romperán en el proceso. Los resultados obtenidos mostraron que CCM2 no era capaz de interaccionar, al menos de manera directa, con ni ERK1/2 ni con MEK o cRAF (**Figura R21B**). En cambio, si se observó la interacción directa con la proteína CCM1 (**Figura R21B**), demostrando que en ESCs esta interacción se encuentra conservada con respecto a lo descrito en la literatura en otros contextos celulares, tanto en especie murina como humana[171,177,191].

Descartada la posibilidad de que CCM2 intervenga de manera directa en la activación de ERK1/2, decidí analizar el estado transcripcional de factores reguladores de esta vía. Para ello empleé dos listas de ontología génica constituidas por genes que codifican para reguladores negativos y positivos (GO:0070373 y GO:0070374, respectivamente) de la activación de ERK1/2 previamente descritos en otros contextos celulares. De manera destacable, se observó una amplia variedad de estos transcritos afectados, entre los que destaca la activación de 11 genes que codifican reguladores negativos y la represión de 21 genes que codifican reguladores positivos de ERK1/2 (**Figura R22A**). Destacar que se observó una marcada reducción en la expresión de *bRaf* mientras que los niveles de *cRaf* y *aRaf* se mantuvieron sin cambios en las células KO para *Ccm2* (**Figura R22B**). Además, se vio como varias citoquinas (p. ej. *Fgf15*, *Apoe* y V*egfb*) que constituyen ligandos activadores de los receptores de membrana en la señalización mediada por ERK1/2 redujeron sus niveles de expresión a nivel de ARN (**Figura R22C**). Cabe destacar también, que *Vegfb* es especialmente relevante en el control de la angiogénesis, y su receptor VEGFR2, codificado por el gen *Flk1*, interacciona directamente con el complejo CCM en las células endoteliales[188]. Por último, la expresión de varios inhibidores de la vía de ERK1/2 (p. ej. *Dusp9*, *Dusp6* y *Epha4*) se encontró ligeramente incrementada significativamente (**Figura R22D**). En concreto, *Dusp9* es una de las principales fosfatasas

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

duales en ESCs, capaz de defosforilar e inactivar directamente a ERK1/2[130].

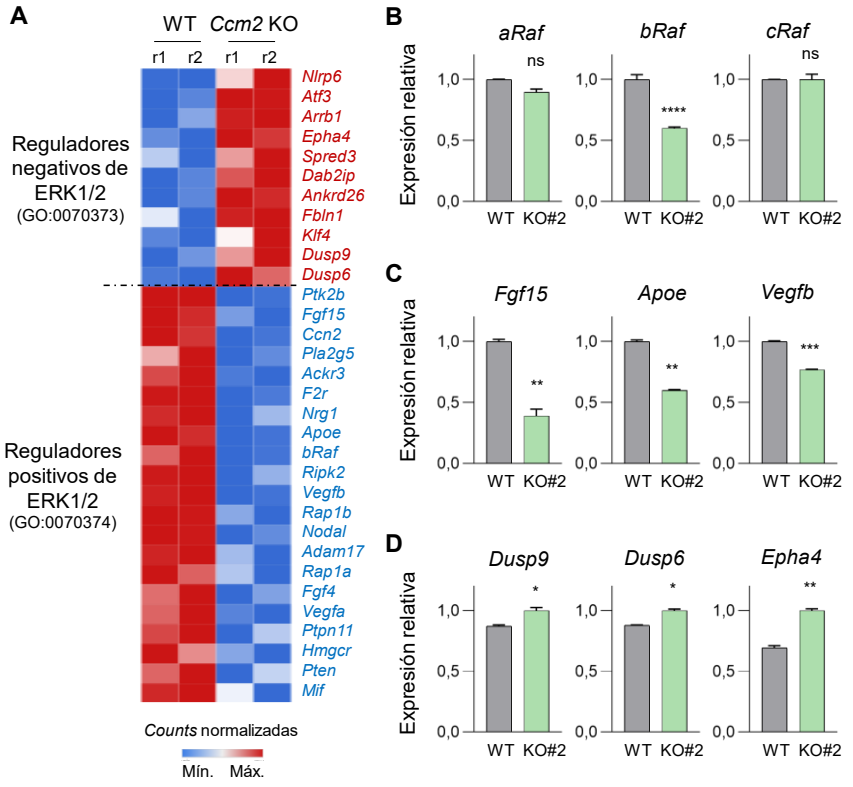


Figura R22. La pérdida de función de CCM2 en ESCs altera los niveles transcripcionales de reguladores positivos y negativos que participan en el control de la activación de ERK1/2. A) Representación gráfica mediante mapa de calor de la expresión de 11 genes que codifican reguladores negativos de ERK1/2 (GO:0070373) y 21 genes que codifican reguladores positivos de esta vía de señalización (GO:0070374), que se encuentran diferencialmente expresados en ESCs KO para *Ccm2* con relación a ESCs WT. Se emplearon los datos de *counts* normalizadas de los RNA-seq generados en este estudio. Representación mediante gráficas de barras de la expresión relativa de las isoformas de *Raf* **(B)**, de 3 citoquinas activadoras de la vía de ERK1/2 **(C)** y de 3 represores de esta vía de señalización **(D)** en ESCs WT y KO para *Ccm2* calculada por qPCR. En **B**, **C** y **D** los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre KO y WT. *p<0,05; **p<0,01; ***p<0,001; ****p<0,0001; ns: no significativo.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

En conjunto, mis datos sugieren que CCM2, al menos de manera indirecta, resulta esencial para la activación de ERK1/2 en ESCs. En línea con esto, la ausencia de CCM2 provoca el incremento y reducción de los niveles de expresión de diversos reguladores negativos y positivos de la vía de ERK1/2, respectivamente.

5.2.4 La inhibición de ERK1/2 contribuye a la desregulación transcripcional causada por la pérdida de CCM2

El estado pluripotente *naïve* se puede estabilizar mediante el cultivo de las ESCs en medio sin suero y la presencia de la citoquina LIF junto con los inhibidores de las quinasas GSK3β y MEK (CHIR99021 y PD0325901, respectivamente)[17,29]. Tras comprobar que CCM2 es crítico para la activación de ERK1/2, me propuse evaluar los efectos a nivel de expresión génica del uso individual de estos dos inhibidores para intentar determinar qué se esconde tras el fenotipo similar a pluripotencia *naïve* observado en las ESCs KO para *Ccm2*. Con este fin, tomé ventaja de estudios previos en los que se analizan diferentes combinaciones de medios de cultivo celular incluyendo presencia conjunta o individual de los inhibidores de GSK3β y MEK[283,284]. En concreto, estos estudios contienen conjuntos de datos de RNA-seq de ESCs derivadas de dos cepas murinas con fondo genético distinto (129 y BL6) cultivadas en: i) condiciones estándar sin presencia de inhibidores (SL), con presencia de los inhibidores de MEK y GSK3β (SL+2i) o sólo presencia del inhibidor de MEK (SL+PD); ii) condiciones *naïve* en ausencia de suero (N2B27) con presencia de los dos inhibidores y LIF (2iL), sin LIF (2i) o con sólo la presencia individual de cada uno de los inhibidores con LIF (PD-L y CHIR-L). En primer lugar, reanalicé *in silico* estos conjuntos de datos de secuenciación masiva con el fin de normalizar la expresión de cada condición contra las ESCs WT cultivadas en condiciones estándar SL. Luego, empleé los valores de cambio transcripcional (FC) en cada condición sobre los 1.363 genes diferencialmente expresados en las ESCs *Ccm2* KO (**Figura R7A**) para generar una matriz de similitud

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

mediante el coeficiente de correlación de Pearson, sometiendo todas las condiciones a una agrupación jerárquica (**Figura R23A**). Así, se observó que, a pesar de existir un sesgo por estudio y tipo celular, las condiciones más cercanas a las ESCs KO para *Ccm2* fueron las que contienen en el medio de cultivo el inhibidor de MEK: SL-PD (129), N2B27-2iL, N2B27-PD-L (129) y N2B27-PD-L (BL6) (**Figura R23A**). Por otro lado, se vio que, aunque tanto la inhibición de GSK3β como la de MEK conllevaban una tendencia de reducción en los niveles de expresión de *Ccm2*, solo la inhibición de ERK1/2 a través de la presencia del inhibidor de MEK (PD0325901) resultó en una disminución significativa (FDR<0,05) en la activación de este gen (**Figura R23B**). Conforme a los resultados *in silico* observados, el cultivo de las células KO para *Ccm2* en medio sin suero con LIF (N2B27-LIF) y en presencia únicamente del inhibidor de GSK3β resultó suficiente para el cambio en la morfología de las colonias ESCs, volviéndolas compactas y regulares de manera similar al medio *naïve* completo (**Figura R23C**).

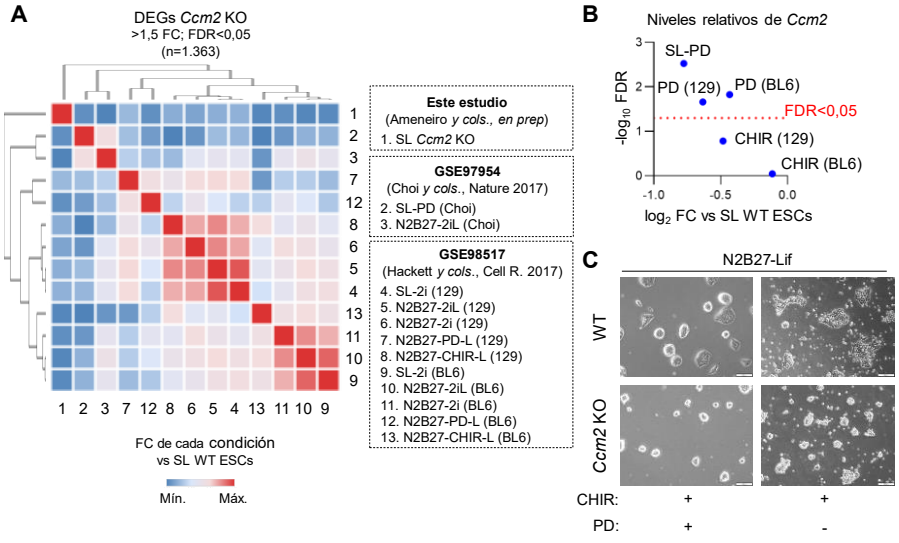


Figura R23. Análisis de similitud entre *Ccm2* KO y la inhibición de MEK o GSK3β en ESCs. A) Representación mediante mapa de calor del análisis de similitud a partir de los valores de cambio (FC) de la expresión de los genes desregulados tras la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

eliminación de CCM2 entre ESCs KO para *Ccm2* en condiciones estándar de cultivo (SL) y ESCs WT en las condiciones de cultivo indicadas. Se empleó el coeficiente de correlación de Pearson para generar la matriz de similitud acoplada a una agrupación jerárquica. Todos los FC fueron calculados contra ESCs WT en SL a partir de datos de RNA-seq generados en esta tesis y en los estudios de Choi y colaboradores (GSE97954)[283] y Hackett y colaboradores (GSE98517)[284]. Las diferentes condiciones de cultivo abarcan la inhibición independiente de MEK (PD) y GSK3β (CHIR) o conjunta (2i) en medio SL o medio N2B27-L en distintas líneas de ESCs. **B)** Representación mediante gráfica de dispersión de los niveles de expresión relativos de *Ccm2* tras la inhibición de MEK (PD) o GSK3β (CHIR). Se representa el umbral de significancia estadística (FDR<0,05). **C)** Imágenes representativas de ESCs WT y KO para *Ccm2* cultivadas en medio sin suero N2B27-LIF y presencia de ambos inhibidores (CHIR y PD) o solo CHIR tras 48h. Las barras representan 100 µm.

Estos resultados revelan que el fenotipo similar al estado pluripotente *naïve* que presentan las ESCs KO para *Ccm2* está mediado en parte por la drástica reducción en los niveles de activación de ERK1/2 que resultan de la pérdida de función de CCM2. Teniendo en cuenta que previamente se ha mostrado que ERK1/2 puede actuar de manera directa en el control transcripcional de genes relacionados con la salida de pluripotencia[105], lo siguiente que me propuse fue valorar los niveles de expresión de las dianas de ERK1/2 a nivel de cromatina en ausencia de función de CCM2. Para este fin tomé ventaja de las dianas de ERK2 identificadas a nivel genómico por el grupo del Dr. Danny Reinberg mediante ChIP-seq en ESCs murinas cultivadas en condiciones estándar (SL)[105]. Mediante un análisis de enriquecimiento de conjuntos de genes (GSEA) utilizando mis datos de RNA-seq en ESCs KO para *Ccm2* y ESCs WT, se observa que las dianas transcripcionales de ERK2 se encontraban significativamente alteradas en ausencia de CCM2 (FDR=0,0 y NES=-1,54) (**Figura R24A**). En particular, el 11,9% (63 de 528 genes) de las dianas de ERK2 se encontraban dentro de los genes con expresión desregulada en más de un 50% significativamente (FC>1,5 y FDR<0,05) en las ESCs mutantes para *Ccm2*, de los cuales 41 se hallaban activados y 22 reprimidos (**Figura R24B**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

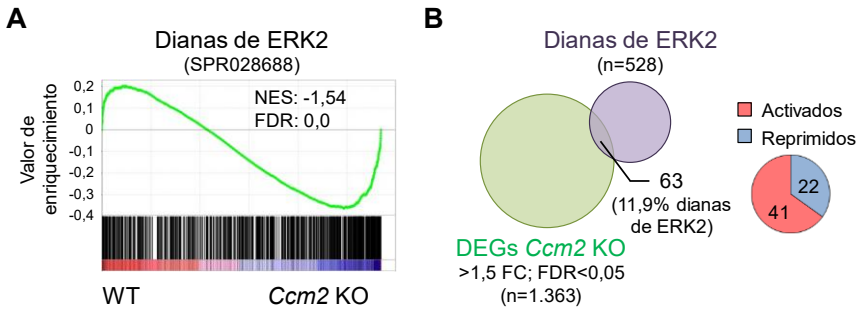


Figura R24. La ausencia de CCM2 afecta a la expresión de dianas transcripcionales de ERK2. A) Análisis de GSEA de los datos de RNA-seq de ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo contra la lista de dianas transcripcionales de ERK2 determinadas en el estudio de Tee y colaboradores (SPR028688)[105]. Se muestran los valores de la puntuación de enriquecimiento normalizada (NES) y de la tasa de descubrimientos falsos (FDR). **B)** (Izquierda) Representación mediante diagrama de Venn de las dianas de ERK2 y los genes diferencialmente expresados en ausencia de CCM2 (FC>1,5, FDR<0,05). (Derecha) Representación gráfica de las 63 dianas desreguladas en las ESCs KO para *Ccm2* indicando el número de éstas que se encuentran activadas o reprimidas transcripcionalmente.

Por otro lado, se ha mostrado que numerosas proteínas, entre las que se encuentran factores de transcripción, son sustratos de la actividad quinasa de ERK1/2 e intervienen en el control de la expresión génica en diversos contextos celulares[285]. Tomando en consideración que sólo el 11,9% de los genes desregulados de manera significativa (FDR<0,05) y con más de un 50% de cambio (FC>1,5) en las ESCs mutantes para *Ccm2* son dianas de ERK2, lo siguiente que me propuse fue determinar qué porcentaje del transcriptoma afectado por la ausencia de CCM2 (1.363 genes desregulados con FC>1,5 y FDR<0,05) correlaciona con la inhibición de la activación ERK1/2 observada en las células *Ccm2* KO. Para ello, comparé estas alteraciones transcripcionales promovidas por la ausencia de CCM2 con las de los diferentes estudios dependientes únicamente de la presencia del inhibidor de MEK (PD0325901) en el cultivo (**Figura R25A**). Así, se observa que el 57% (n=777) de los genes desregulados en las ESCs KO para *Ccm2* también muestran su expresión

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

significativamente alterada (FC>1,5 y FDR<0,05) en alguna de las tres condiciones analizadas en las que se encuentra presente el inhibidor de MEK (SL-PD, N2B27-PD-L 129 o N2B27-PD-L BL6) (**Figura R25A**). Un análisis más en profundidad mostró la existencia de cuatro clases de genes en función de los cambios transcripcionales detectados con respecto a la pérdida de función de CCM2: i) la clase 1, que comprende 221 genes, representa aquellos genes cuya desregulación sigue la misma tendencia en las tres condiciones de cultivo indicadas y mis ESCs KO para *Ccm2*; ii) la clase 2, compuesta por 219 genes que presentan activación o represión en *Ccm2* KO y en dos de las condiciones de cultivo analizadas; iii) la clase 3, con 192 genes cuya tendencia en los cambios de expresión es similar entre la ausencia de CCM2 y en una de las condiciones de cultivo; y iv) la clase 4, constituida por 145 genes que muestran una tendencia opuesta entre *Ccm2* KO y el resto de las condiciones con presencia del inhibidor de MEK (**Figura R25B**). Seguidamente, comprobé cuántos de los 777 genes son dianas de ERK2[105]. En línea con los resultados anteriores (**Figura R24B**), se observó que, en cada una de las cuatro clases, la desregulación de la gran mayoría de los genes que se encontraban activados o reprimidos en ausencia de CCM2 (un mínimo del 80%) no depende de manera directa de la unión de ERK2 en la cromatina para su control transcripcional (**Figura R25C**).

En conjunto, mis datos apoyan que el fenotipo *naïve* observado tras la eliminación de CCM2 se debe en gran parte a la deficiente activación de ERK1/2. No solo el análisis jerárquico muestra una mayor similitud transcripcional entre *Ccm2* KO y células cultivadas en presencia de inhibición de MEK y no de GSK3β, sino que la presencia del inhibidor PD0325901 en el medio es suficiente para reducir significativamente el nivel de expresión de *Ccm2*. Además, mis resultados muestran que la inhibición química de MEK comparte con las ESCs KO para *Ccm2* la desregulación de 777 genes (57% de los cambios transcripcionales sin CCM2), de los cuales solo el 18,7% (n=145) muestran una tendencia opuesta.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

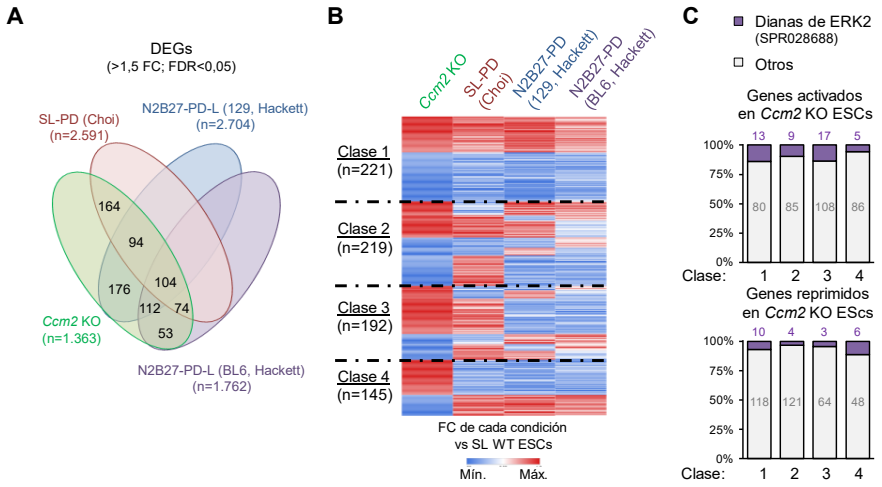


Figura R25. La inhibición de ERK1/2 contribuye a gran parte de la desregulación transcripcional mediada por la ausencia de CCM2. A) Representación mediante diagrama de Venn de los genes diferencialmente expresados (DEGs) en las ESCs KO para *Ccm2* en condiciones estándar de cultivo (SL) y en las ESCs cultivadas en SL o en N2B27-L con presencia del inhibidor de MEK (PD). Se muestran los 777 genes comunes entre *Ccm2* KO y al menos una de las otras condiciones indicadas. Todos los cambios de expresión (FC) fueron calculados mediante la normalización con sus respectivos controles (ESCs WT en SL). **B)** Representación mediante mapa de calor empleando los valores de cambio de expresión (FC) de los 777 genes analizados. Se representan los genes con la misma tendencia de desregulación en todas las condiciones (Clase 1), en tres de ellas (Clase 2), en dos de ellas (Clase 3) y los genes con desregulación opuesta en *Ccm2* KO (Clase 4). **C)** Representación mediante gráfica de barras del porcentaje de genes activados o reprimidos pertenecientes a cada clase. Se indica el número de genes dianas de ERK2 y aquellos en los que no se detecta la presencia de esta quinasa en sus regiones próximas reguladoras a nivel de cromatina.

5.2.5 PRC2 participa en los cambios transcripcionales tras la pérdida de CCM2

Tras determinar que ERK1/2 participa de manera directa e indirecta en la regulación de la expresión génica en la que interviene CCM2 en ESCs, pero que no explica en su totalidad el paisaje transcripcional observado en las ESCs KO para *Ccm2*, me propuse analizar que otros reguladores transcripcionales podrían estar

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

involucrados. Mediante el empleo de la herramienta para el análisis de enriquecimiento ChEA3[255], diseñada para comprobar si una lista de genes se encuentra enriquecida en transcritos identificados como objetivos putativos de factores de transcripción y reguladores epigenéticos a partir de la integración de datos procedentes de ChIP-seqs publicados, pude determinar qué factores se unen con más probabilidad en la cromatina de los 1.363 genes desregulados (FC>1,5 y FDR<0,05) sin CCM2. Entre los reguladores transcripcionales con más enriquecimiento en los genes desregulados en ausencia de CCM2 se encontraban varias proteínas del grupo *Polycomb* (PcG) (**Figura R26A**), que se corresponden a represores epigenéticos que intervienen en el control de la expresión de genes esenciales en desarrollo embrionario y diferenciación celular[286]. En concreto SUZ12, MTF2 y EZH2 pertenecen al Complejo Represor *Polycomb* 2 (PRC2), y RNF2 al Complejo Represor *Polycomb* 1 (PRC1). Siendo las proteínas de PRC2 las más representadas, analicé el estado transcripcional de las dianas de SUZ12 en ausencia de CCM2 tomando ventaja de un ChIP-seq publicado previamente en ESCs[111]. A través de un análisis GSEA se observó que dichas dianas presentaban su expresión alterada significativamente (NES=-1,94; FDR=0,0) en las ESCs KO para *Ccm2* (**Figura R26B**). De los 1.363 genes desregulados sin CCM2, el 34,8% (n=475) son dianas de SUZ12 (**Figura R26C**). Asimismo, el análisis de la expresión de estos 475 genes que se encuentran unidos por SUZ12 mostró que la ausencia de CCM2 conllevaba tanto su activación (73,4%, n=349) como su represión (26,5%, n=126) (**Figura R26C**). Además, el análisis de las vías de señalización KEGG que se ven afectadas significativamente por la desregulación de estos 475 genes dianas de SUZ12 reveló la presencia de genes que intervienen en la regulación de la vía de las MAPK, la regulación de las ESCs y otras vías relacionadas con procesos importantes tanto en pluripotencia como en la biología endotelial (**Figura R26D**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

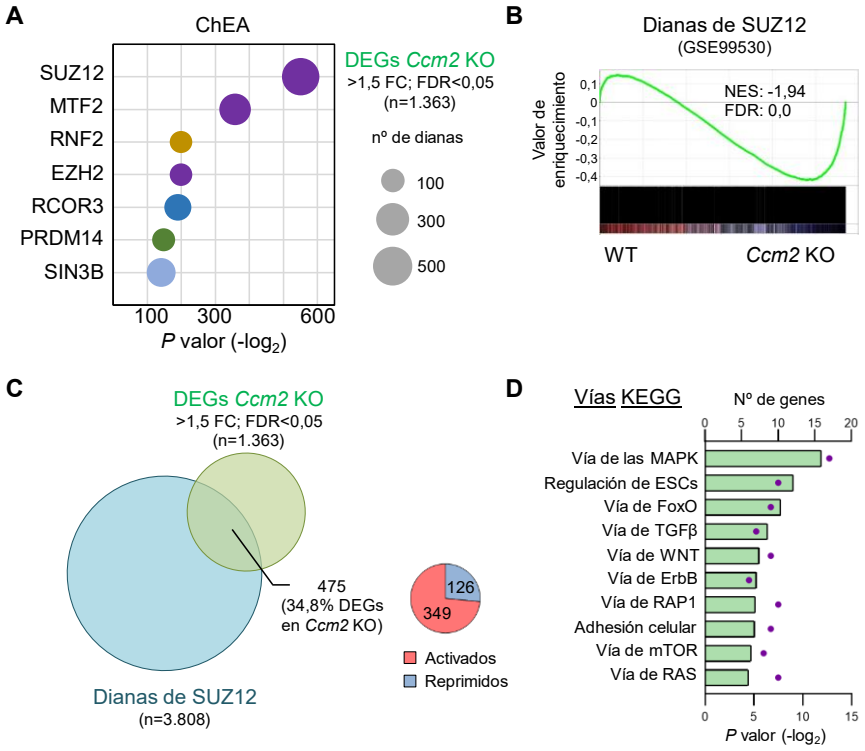


Figura R26. La ausencia de CCM2 altera la expresión de genes diana de PRC2. A) Representación gráfica del análisis de enriquecimiento mediante la herramienta ChEA3 empleando los 1.363 genes diferencialmente expresados tras la eliminación de *Ccm2*. Se muestra la probabilidad de ser dianas de diversos reguladores de la transcripción a través de la significancia y el número de dianas putativas entre los genes desregulados. **B)** Análisis GSEA de los datos de RNA-seq de ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo comparado con la lista de dianas de SUZ12 determinadas en el estudio de Mas y colaboradores (GSE99530)[111]. Se muestra el valor de NES y FDR. **C)** (Izquierda) Representación mediante diagrama de Venn de las dianas de SUZ12 y los genes diferencialmente expresados en ausencia de CCM2. (Derecha) Representación gráfica de las 475 dianas desreguladas indicando el número de genes activados o reprimidos en ausencia de CCM2. **D)** Análisis de las vías de señalización KEGG empleando los 475 genes dianas de SUZ12 y desregulados tras la eliminación de *Ccm2*. Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de genes involucrados en cada vía. Las barras representan el *P* valor y los puntos el número de genes. En **A** y **D** los términos representados se seleccionaron entre los más enriquecidos.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tras comprobar que un gran porcentaje de los genes desregulados en las ESCs KO para *Ccm2* son dianas de PRC2, decidí analizar si la pérdida de función de CCM2 podría afectar a los niveles proteicos de componentes de este complejo regulador epigenético. De modo destacado, se observó que las células mutantes para *Ccm2* presentaban un incremento de los niveles proteicos a nivel global de SUZ12 y EZH2 comparado con las ESCs WT (**Figura R27A**).

Tal y como se describió en la introducción (sección 2.2.2.2), los complejos PRC1 y PRC2 son los encargados del depósito de un grupo ubiquitina en la lisina 119 de la histona H2A (H2AK119ub) y de la metilación de la lisina 27 de la histona H3 (H3K27me3), respectivamente[83,84]. Por esta razón, me propuse evaluar los niveles globales de estas dos marcas en extractos de la fracción proteica unida a cromatina obtenida de ESCs WT y KO para *Ccm2*. Mientras que no se apreciaron alteraciones en los niveles globales de la marca H2AK119ub, la ausencia de CCM2 conllevó un aumento considerable de los niveles globales de la modificación química H3K27me3 depositada por PRC2 en la cromatina (**Figura R27B**). A este respecto, cabe destacar que el aumento detectado tanto en los niveles globales proteicos de los componentes centrales del complejo PCR2 como de su marca asociada H3K27me3 en las ESCs KO para *Ccm2* es consistente con lo mostrado previamente en la literatura para ESCs cultivadas en medio *naïve* (2i) en comparación con las ESCs cultivadas en medio estándar (SL)[103].

Por otro lado, en las células pluripotentes existe un grupo de genes denominados bivalentes o genes asociados al desarrollo, caracterizados por la coexistencia en sus regiones promotoras de las marcas H3K27me3 y H3K4me3[98]. Teniendo en cuenta la alteración observada en los niveles globales de H3K27me3 en ausencia de CCM2, procedía a evaluar si las ESCs KO para *Ccm2* mostraban alteraciones transcripcionales en los genes bivalentes que se encuentran normalmente silenciados o poco expresados en ESCs. Tomando ventaja de un estudio previo en el que se identifica la existencia de 3.868 genes

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

bivalentes en ESCs cultivadas en condiciones estándar (SL)[111], y mediante un análisis GSEA, se observó que las células KO para *Ccm2* presentaban enriquecimiento de estos transcritos (NES=-1,99; FDR=0,0) (**Figura R27C**).

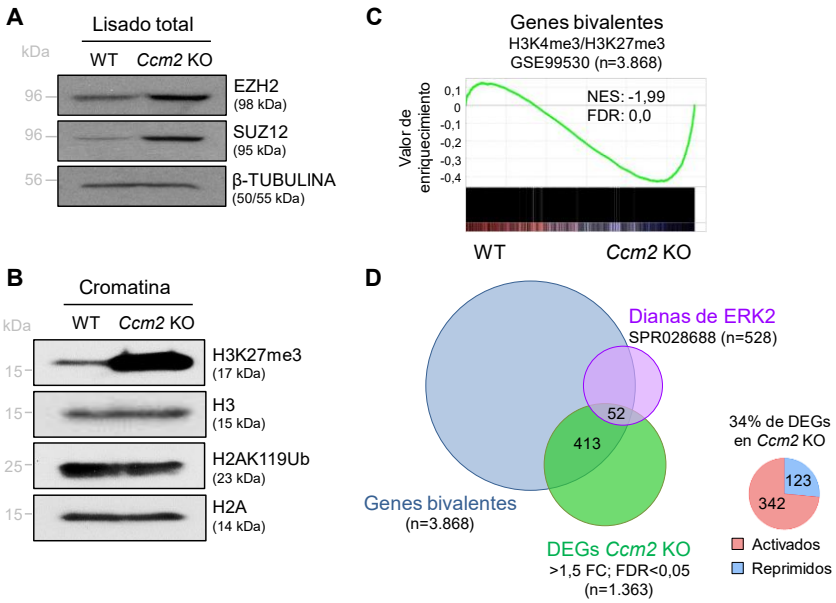


Figura R27. CCM2 participa en la regulación transcripcional de genes bivalentes. **A)** Western blot para detectar los niveles proteicos globales de componentes de PRC2 (EZH2 y SUZ12) en ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo. ß-TUBULINA se empleó como control de carga. **B)** Western blot para detectar los niveles globales de las marcas depositadas por PRC2 (H3K27me3) y PRC1 (H2AK119Ub) en la cromatina de ESCs WT o KO para *Ccm2* en condiciones estándar de cultivo. H3 y H2A se emplearon como controles de carga. **C)** Análisis GSEA de los datos de RNA-seq de ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo contra la lista de genes bivalentes (presentan H3K4me3 y H3K27me3) determinados en el estudio de Mas y colaboradores (GSE99530)[111]. Se muestra el valor de NES y FDR. **D)** (Izquierda) Representación mediante diagrama de Venn de los genes bivalentes, las dianas de ERK2 y los genes diferencialmente expresados en ausencia de CCM2. (Derecha) Representación gráfica de los genes comunes indicando el número que se activa o reprime.

Un análisis más exhaustivo mostró que el 34,1% (n=465) de los genes desregulados (FC>1,5 y FDR<0,05) tras la eliminación de CCM2

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

constituyen genes bivalentes (**Figura R27D**). Además, se pudo apreciar que CCM2 contribuye a la represión del 73,5% (n=342) de estos genes (**Figura R27D**). Dado que mis datos muestran que CCM2 es esencial para la activación de ERK2 en las ESCs (**Figura R20**) y que se ha mostrado que la unión de ERK2 en su forma activa interviene en la regulación de ciertos genes bivalentes promoviendo un estado de transcripción silenciado preparado para la expresión (*poised*)[105,151], decidí analizar si ERK2 podría participar de manera directa en la cromatina en la regulación transcripcional de los 465 genes bivalentes cuya expresión se encuentra alterada significativamente en ausencia de CCM2. A este respecto, aunque el 71,9% (380 de 528 genes) de las dianas de ERK2 poseen en sus regiones promotoras la presencia de H3K4me3 y H3K27me3, sólo el 9,8% de éstos (52 de 528 genes) se encontraban significativamente alterados sus niveles en más de un 50% en las ESCs KO para *Ccm2* (**Figura R27D**). Estos resultados sugieren que la gran mayoría de genes bivalentes (413 de 465 genes), en los que CCM2 participa en el control de su expresión transcripcional, no dependen de la presencia de ERK2 en sus regiones reguladoras.

Tras haber identificado una importante correlación entre la presencia a nivel genómico de PRC2 y genes cuya expresión se encuentra alterada en ausencia de CCM2, lo siguiente que me propuse fue analizar si la actividad catalítica de este complejo represor epigenético podría contribuir a las funciones regulatorias de la expresión génica en las que interviene CCM2 en ESCs. Para ello, tomé ventaja de la disponibilidad de un estudio previo en el que analizaron mediante RNA-seq el efecto de la ausencia de EZH2, el componente de PRC2 con actividad catalítica, en ESCs cultivadas en condiciones estándar (SL)[287]. Mediante la comparación de los perfiles transcripcionales de las ESCs *Ccm2* KO y las ESCs *Ezh2* KO se identificaron 523 genes comunes desregulados de manera significativa (FC>1,5 y FDR<0,05), que representaban el 38,4% del total de genes desregulados sin CCM2 (**Figura R28A**). Dado que las células KO para *Ccm2* mostraban un aumento global de la actividad catalítica de PRC2

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

y de los niveles proteicos de sus componentes, incluido EZH2 (**Figuras R27A** y **R27B**), cabría esperar que los genes desregulados comunes mostrasen una tendencia opuesta en sus patrones de expresión. De manera consecuente, se observó que el 73,4% (384 de 523 genes) de estos genes presentaban perfiles transcripcionales opuestos entre las ESCs KO para *Ccm2* y las ESCs KO para *Ezh2* (**Figura R28B**). Este grupo de 384 genes está compuesto por las clases 1 y 2, correspondientes a 135 genes reprimidos (p. ej. *Epcam*, *Dnmt3l*, *Dnmt3b*, *Bmi1*, *Ehmt2* y *Bend3*) y 249 genes activados (p. ej. *Fn1*, *Tfap2c*, *Kit*, *Bmp7*, *Eya1* y *Slit2*) en el KO de *Ccm2* (**Figura R28B**). Luego, analicé con más detalle cada una de estas dos clases de genes en función de si presentaban marcas bivalentes H3K4me3 y H3K27me3 (K4K27) o si eran dianas de ERK2. Con respecto a la clase 1 se observó que la constituyen 109 genes (80,7%) que no presentaban ninguna de estas características (p. ej. *Epcam*, *Dnmt3l*, *Ehmt2* y *Bend3*) y 21 genes (15,6%) que eran bivalentes (p. ej. *Dnmt3b*), de los cuales 6 también eran dianas de ERK2 (p. ej. *Bmil*) (**Figura R28C**). En cuanto a los genes pertenecientes a la clase 2, 114 (45,8%) se clasificaban como bivalentes (p. ej. Kit) y en 14 de estos (5,6%) se encontraba presente ERK2 (p. ej. *Tfap2c*, *Bmp7* y *Slit2*) (**Figura R28C**). Mediante análisis de ontología génica de los procesos biológicos asociados a cada una de estas clases se identificó que: i) los transcritos reprimidos tras la ausencia de CCM2 (clase 1) se encontraban asociados con procesos de metilación del ADN, destacando los genes *Dnmt3l*, *Dnmt3b*, *Bmi1*, *Ehmt2* y *Bend3*, y procesos relacionados con desarrollo embrionario y la diferenciación de células troncales, a través de genes como *Epcam*; ii) los transcritos incrementados en las ESCs KO para *Ccm2* (clase 2) se encontraban relacionados de manera significativa con procesos de diferenciación celular, destacando genes como *Tfap2c*, *Kit*, *Bmp7*, *Eya1* y *Slit2*, o angiogénesis a través de genes como *Fn1* (**Figura R28D**). Cabe resaltar que estos 6 transcritos incrementados en la clase 2 presentan sus niveles de expresión más elevados en el estado pluripotente *naïve* comparados con otros estados alternativos de pluripotencia.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

El conjunto de estos resultados revela que la ausencia de CCM2 provoca una alteración en la regulación de la expresión génica dependiente del complejo PRC2 que contribuye a parte de la desregulación transcripcional detectada en las ESCs KO para *Ccm2*.

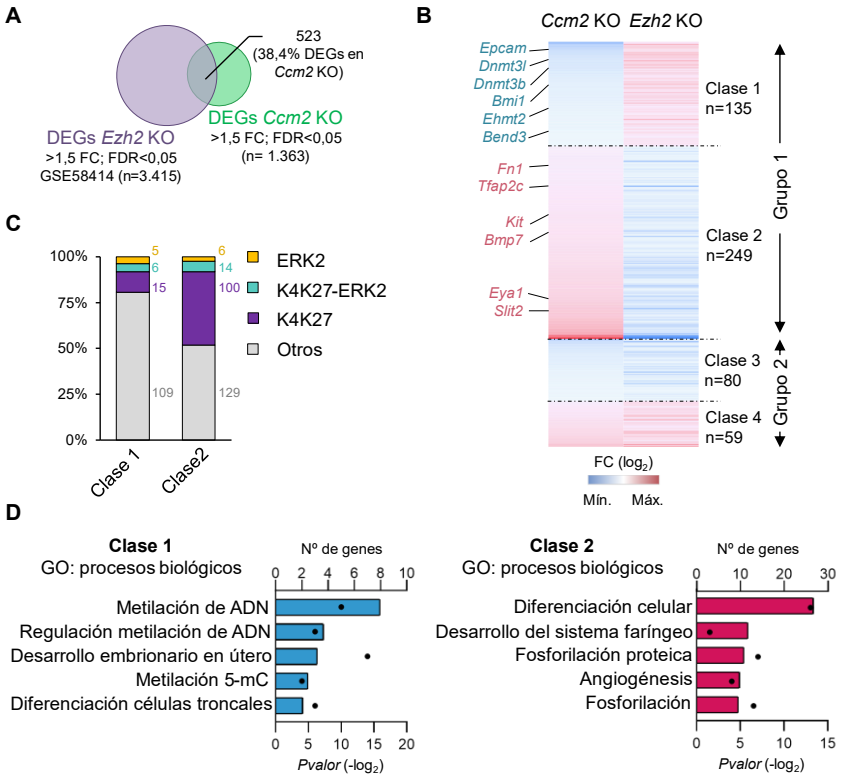


Figura R28. La ausencia de actividad catalítica de PRC2 explica parte de la desregulación transcripcional de las ESCs KO para *Ccm2*. A) Representación mediante diagrama de Venn de los genes diferencialmente expresados en las ESCs KO para *Ccm2* y para *Ezh2* en condiciones estándar de cultivo mostrando los genes comunes en las dos condiciones. Se emplearon los datos de RNA-seq generados en esta tesis y en el estudio de Das y colaboradores (GSE58414)[287]. **B)** Representación mediante mapa de calor de los valores de cambio transcipcional (FC>1,5 y FDR<0,05) de los 523 genes comunes. Se representan los genes con tendencia opuesta de desregulación (Grupo 1) y con la misma tendencia (Grupo 2). Dentro de cada grupo, se ordenaron los genes en clases dependiendo de su represión o activación. **C)** Representación mediante gráfica de barras del porcentaje de genes de la clase 1 y 2 que constituyen genes bivalentes (K4K27), genes bivalentes dianas de ERK2 (K4K27-ERK2) o solo dianas de ERK2. Se indica el número de genes en cada condición. **D)**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Análisis de ontología génica (GO) de procesos biológicos empleando los genes de la clase 1 (izquierda) y clase 2 (derecha). Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de genes desregulados involucrados en cada proceso. Las barras representan el *P* valor y los puntos el número de genes. Los términos representados se seleccionaron entre los más enriquecidos. Genes desregulados involucrados en metilación del ADN, diferenciación de células troncales, diferenciación celular y angiogénesis se resaltan en el panel **B**.

5.2.6 Las *Ccm2* KO presentan niveles reducidos de metilación en ADN

Teniendo en cuenta que: i) los resultados anteriores muestran que CCM2 y PRC2 participan en el control transcripcional de genes relacionados con la regulación del estado de metilación en ADN (**Figura R28**), ii) la hipometilación del ADN a nivel global es una característica destacada de la pluripotencia *naïve*[69,104] y iii) la metilación en ADN gobierna la progresión durante desarrollo desde *naïve* hacia el estado celular *primed*[73,288], decidí evaluar cómo CCM2 podría afectar a esta importante reconfiguración epigenómica en ESCs. Con este fin, en primer lugar, realicé un análisis mediante GSEA de una lista de genes involucrados en la metilación del ADN (GO:0006306), encontrando que estos transcritos se encuentran enriquecidos (NES=1,587; FDR=0,008) en las ESCs WT en detrimento de las células KO para *Ccm2* (**Figura R29A**). Luego, decidí analizar los niveles globales de metilación de ADN en las ESCs WT y KO para *Ccm2*. Como se describió en la introducción (sección 2.2.2.1), el ciclo de metilación del ADN no depende únicamente de la deposición correcta de 5-metilcitosina (5-mC), sino del proceso de desmetilación activa, por el que se genera 5-hidroximetilcitosina (5-hmC) y los sucesivos intermediarios de oxidación[59,62-64,67]. Por esta razón, a partir de ADN total purificado de ESCs cultivadas en medio estándar (SL) determiné los niveles globales de 5-mC y 5-hmC en el ADN en presencia y ausencia de CCM2 mediante ensayos cuantitativos de inmunoadsorción ligado a enzima (ELISA) específicos para estas dos marcas epigenéticas (**Figura R29B**). Los resultados obtenidos mostraron una reducción significativa de los niveles globales de 5-mC en ADN en las ESCs KO

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

para *Ccm2* comparado con las ESCs WT (**Figura R29C**). Además, también se detectó una reducción en los niveles de 5-hmC en ausencia de CCM2, lo que podría ser consecuencia de menor abundancia en los niveles de 5-mC en las ESCs KO para *Ccm2* (**Figura R29D**), y se encontraría en línea con un estudio reciente en el que muestran por espectrometría de masas que las ESCs cultivadas en medio *naïve* (2iL) presentan niveles más reducidos de 5-hmC en comparación con las ESCs cultivadas en medio estándar (SL)[103].

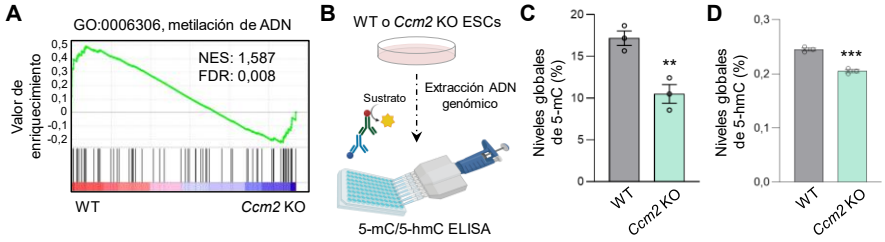


Figura R29. La ausencia de CCM2 altera la abundancia de 5-mC y 5-hmC en ADN. **A)** Análisis GSEA de los datos de RNA-seq de ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo contra la lista de genes involucrados en la metilación del ADN (GO:0006306). Se muestra el valor de NES y FDR. **B)** Representación esquemática de la estrategia seguida para analizar los niveles globales de 5-mC y 5-hmC. Se extrajo ADN genómico de ESCs WT o KO para *Ccm2* en condiciones estándar de cultivo (SL) y se cuantificaron los niveles de 5-mC y 5-hmC mediante ensayo de inmunoadsorción ligado a enzima (ELISA). **C)** Representación mediante gráfica de barras del porcentaje detectado de citosinas metiladas (5-mC) en el ADN de ESCs WT o KO para *Ccm2* cultivadas en SL. **D)** Representación mediante gráfica de barras de la abundancia en el genoma de citosinas hidroximetiladas (5-hmC) en presencia o ausencia de CCM2 en ESCs. En **C** y **D** n=3 experimentos independientes. Los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre KO y WT. **p<0,01; ***p<0,001.

5.2.7 La pérdida de CCM2 altera la abundancia de enzimas que gobiernan la metilación genómica

Para conocer más en detalle las razones por las que las ESCs KO para *Ccm2* presentan niveles reducidos de 5-mC y 5-hmC, me propuse analizar cómo se encuentran las proteínas encargadas tanto de depositar los grupos metilo en las citosinas (DNMTs, del inglés *DNA*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

methyltransferase) como las enzimas responsables de la eliminación de manera activa de esta marca epigenética (TETs, del inglés *Ten-eleven translocation*). Primero, evalué mediante qPCR el nivel de expresión de los genes codificantes de las DNMTs que introducen 5-mC y las dioxigenasas de metilcitosina TETs en presencia y ausencia de CCM2 en condiciones estándar de cultivo (SL).

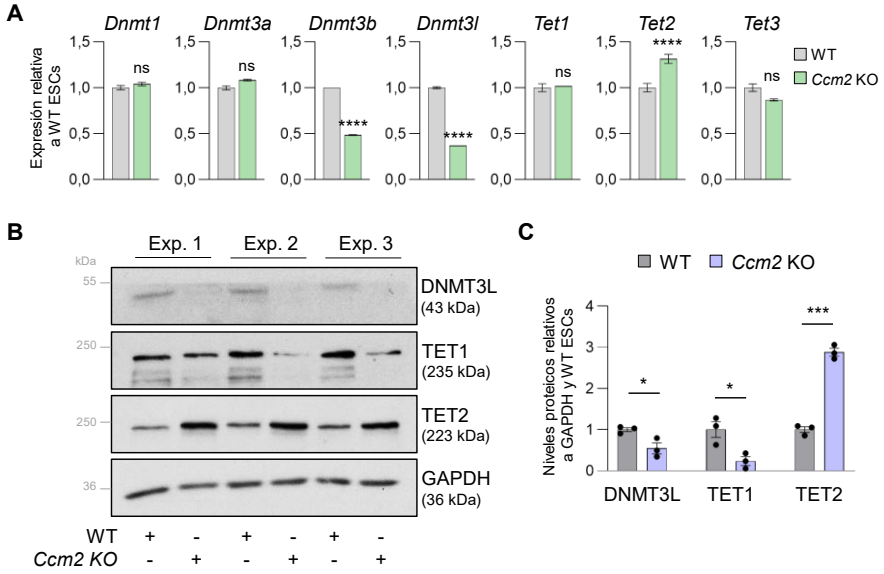


Figura R30. La eliminación de CCM2 modifica los niveles globales de DNMTs y TETs. A) Representación mediante gráfica de barras de la expresión relativa de *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Dnmt3l*, *Tet1*, *Tet2* y *Tet3* en ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo por qPCR. **B)** Western blot para detectar los niveles proteicos globales de DNMT3L, TET1 y TET2 en ESCs WT o KO para *Ccm2* en condiciones estándar de cultivo en tres experimentos independientes. GAPDH se empleó como control de carga. **C)** Cuantificación mediante gráfico de barras de la expresión proteica analizada en **B** con relación a GAPDH. En **A** y **C** los datos se presentan como media ± s.e.m. Se aplicó t-test de Student de dos colas no pareado para el cálculo estadístico entre KO y WT. *p<0,05; **p<0,01; ***p<0,001; ****p<0,0001; ns: no significativo.

De manera consistente con mis datos de RNA-seq (**Figura R28B**), se confirmó que las ESCs KO para *Ccm2* muestran niveles significativamente reducidos tanto del gen *Dnmt3b*, que codifica la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

enzima encargada de la metilación 5-mC *de novo*, como del gen que codifica *Dnmt3l*, que interviene en la modulación de la actividad catalítica de DNMT3A y DNMT3B (**Figura R30A**). En cuanto a los niveles de expresión de los genes que codifican cada una de las tres proteínas TET que existen en eucariotas, se vio que CCM2 reprime de manera significativa a *Tet2* (**Figura R30A**). Segundo, procedí a analizar los niveles de expresión proteica de estas enzimas mediante western blot en las ESCs WT y KO para *Ccm2* en las mismas condiciones de cultivo. Se observó que la abundancia de DNMT3L se encontraba reducida y la de TET2 aumentada significativamente en las ESCs KO comparado con las ESCs WT (**Figuras R30B** y **R30C**), en línea con la expresión transcripcional observada. Con respecto a los niveles proteicos de TET1, se observó que en ausencia de CCM2 se encontraban reducidos (**Figuras R30B** y **R30C**), lo que sugiere posibles mecanismos de regulación a nivel postranscripcional o de traducción, ya que sus niveles de expresión génica no se encontraban afectados en las ESCs mutantes para *Ccm2*.

5.2.8 DNMTs y TETs intervienen en el paisaje transcripcional de las *Ccm2* KO

Tanto en desarrollo embrionario temprano como en cultivo celular, la progresión del estado pluripotente *naïve* al *primed* se encuentra dinámicamente regulada a nivel de expresión génica por la metilación en ADN[69,70,104,288]. En concreto, la deposición de 5-mC por DNMT3B resulta esencial en el control de la pluripotencia *primed* tanto en desarrollo[289] como en cultivo celular[290]. Por otro lado, se ha mostrado que la actividad catalítica de TET1 y TET2 desempeña funciones opuestas en estados alternativos de pluripotencia, siendo la de TET2 necesaria para la pluripotencia *naïve* y la de TET1 para el establecimiento y mantenimiento de la pluripotencia *primed*[73]. En base a estas evidencias, cabe destacar que los niveles reducidos de DNMT3B y TET1 junto con el aumento de TET2 tras la pérdida de función de CCM2 (**Figura R30**), sugiere que estas enzimas se encuentran detrás

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de la regulación del fenotipo pluripotente *naïve* que presentan las ESCs KO para *Ccm2*. Por esta razón decidí analizar el papel que podrían ejercer DNMT3B, TET1 y TET2 en el control del paisaje transcripcional que presentan las ESCs cultivadas en condiciones estándar (SL) en ausencia de CCM2. En primer lugar, me centré en investigar cómo la presencia de CCM2 afecta al nivel de expresión de los genes que son dianas de estas enzimas. Con este fin, analicé mediante GSEA el estado de sus dianas génicas en ESCs cultivadas en las mismas condiciones de SL tomando ventaja de la disponibilidad de ChIP-seqs para cada una de estas proteínas publicados previamente[291-293].

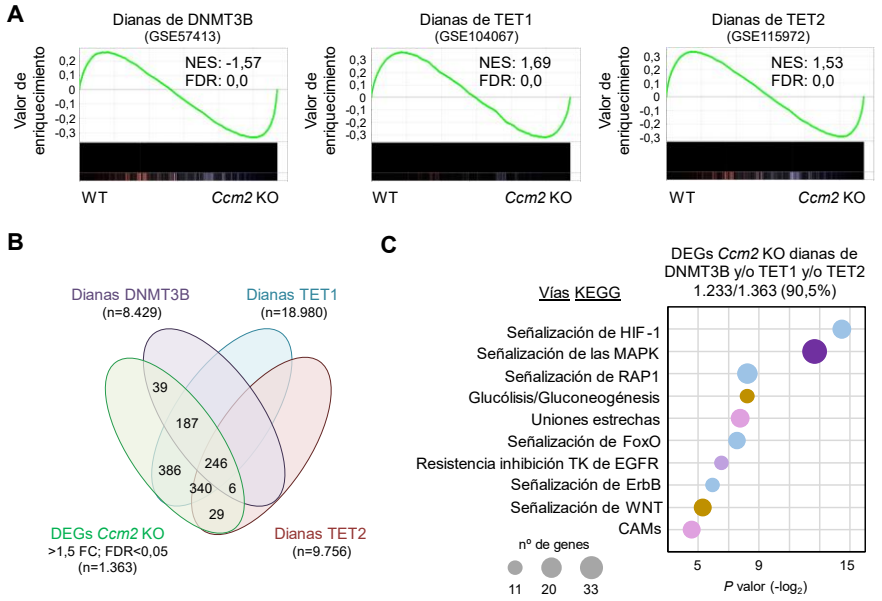


Figura R31. DNMT3B, TET1 y TET2 se unen a genes desregulados transcripcionalmente en ausencia de CCM2. A) Análisis GSEA de los datos de RNA-seq de ESCs WT y KO para *Ccm2* en condiciones estándar de cultivo contra la lista de dianas reanalizadas a partir de los ChIP-seq disponibles para DNMT3B (GSE57413)[291], TET1 (GSE104067)[292] y TET2 (GSE115972)[293]. Se muestra el valor de NES y FDR. **B)** Representación mediante diagrama de Venn de los genes diferencialmente expresados (DEGs) en las ESCs KO para *Ccm2* en condiciones estándar de cultivo y las dianas de DNMT3B, TET1 y TET2. Se muestran 1.233 genes desregulados en ausencia de CCM2 que son dianas al menos de una de las proteínas analizadas. **C)** Análisis de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

las vías de señalización KEGG empleando los 1.233 genes dianas de DNMT3B, TET1 o TET2 y desregulados tras la eliminación de CCM2. Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de genes involucrados en cada vía. Los términos representados se seleccionaron entre los más enriquecidos.

De manera destacada, se observó que los genes unidos por DNMT3B se encontraban enriquecidos en ausencia de CCM2 (NES=-1,57 y FDR=0,0), mientras que las dianas tanto de TET1 (NES=1,69 y FDR=0,0) como de TET2 (NES=1,53 y FDR=0,0) estaban enriquecidas en las ESCs WT (**Figura R31A**). Luego, un análisis comparativo de las dianas de DNMT3B, TET1 y TET2 con los 1.363 genes desregulados (FC>1,5 y FDR<0,05) tras la eliminación de *Ccm2* reveló que el 90,5% de éstos (n=1.233) se encontraban unidos por al menos una de las tres enzimas (**Figura R31B**). Mediante análisis KEGG sobre estos 1.233 genes diana desregulados en las ESCs KO para *Ccm2*, se identificaron entre las vías de señalización significativamente más afectadas aquellas relacionadas con la función endotelial (por ejemplo, la vía de HIF-1, la vía de RAP1, la vía de FoxO o la vía de uniones estrechas) y la vía de las MAPK (**Figura R31C**).

Para entender mejor como DNMT3B, TET1 y TET2 podrían contribuir al control de la expresión de los genes desregulados en las ESCs KO para *Ccm2*, analicé la distribución genómica de estos tres reguladores epigenéticos en los genes activados (n=769) y en los genes reprimidos (n=594) en ausencia de CCM2. Se observó que TET1 y TET2 se encontraban enriquecidos alrededor del sitio de inicio de la transcripción (TSS, del inglés *transcription start site*) tanto en genes que aumentan su expresión como en los que la reducen en las ESCs KO para *Ccm2*, mientras que DNMT3B mostró una distribución menos enriquecida en el TSS de estos genes analizados (**Figura R32A**). De manera destacable, se vio que de los 769 genes regulados al alza (FC>1,5 y FDR<0,05) en ausencia de CCM2, el 40,4% (n=311) eran dianas de DNMT3B, el 85,3% (n=656) de TET1 y el 46,5% (n=358) de TET2 (**Figura R32B**). En cuanto a los 594 genes regulados a la baja (FC>1,5 y FDR<0,05) en las ESCs KO para *Ccm2*, también se observó

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

la presencia de DNMT3B en el 28,1% (n=167) de éstos, TET1 en el 84,7% (n=503) y TET2 en el 44,3% (n=263) (**Figura R32B**).

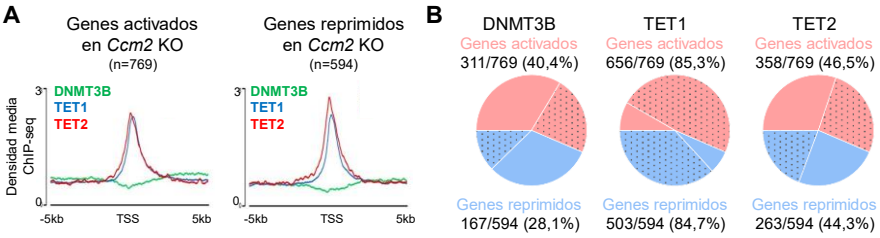


Figura R32. DNMT3B, TET1 y TET2 se encuentran presentes en una fracción de genes desregulados tras la pérdida de CCM2. A) Representación mediante histograma de la densidad media de unión de DNMT3B, TET1 y TET2 a las regiones que flanquean el TSS (± 5 kb) de los 769 genes activados (izquierda) y 594 genes reprimidos (derecha) en ausencia de CCM2. Se emplearon los datos reanalizados a partir de los datos en crudo disponibles en los ChIP-seq en ESCs para DNMT3B (GSE57413)[291], TET1 GSE104067)[292] y TET2 (GSE115972)[293]. **B)** Representación gráfica de los genes activados (rojo) y reprimidos (azul) en *Ccm2* KO indicando el porcentaje y número de genes que constituyen dianas (resaltado con puntos) de DNMT3B (izquierda), de TET1 (centro) o de TET2 (derecha).

Lo siguiente que me propuse fue evaluar el impacto transcripcional dependiente e independiente de la presencia de estos reguladores epigenéticos de manera individual en los 1.363 genes desregulados (FC>1,5 y FDR<0,05) en las ESCs KO para *Ccm2*. Para ello, tomé ventaja de estudios previos en los que analizaron a través de RNA-seq el efecto de la pérdida de función de DNMT3B, TET1 o TET2 en ESCs cultivadas en condiciones estándar (SL)[294,295]. La comparación de las ESCs KO para *Dnmt3b*, *Tet1* y *Tet2* reveló que el 41,7% de los genes desregulados en las ESCs KO de *Ccm2* (n=569) se encontraban también alterados con un cambio en sus niveles de expresión de más del 50% de manera significativa en al menos una de las otras ESCs KO analizadas (**Figura R33A**). En concreto, mientras que la ausencia de DNMT3B y de TET2 resultaba en la desregulación del 77,15% (n=439) y 33,04% (n=188) de estos 569 genes, respectivamente, la pérdida de TET1 sólo afectaba al 5,8% (n=33) de éstos (**Figura R33B**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Finalmente, mediante análisis de las vías KEGG asociadas con estos 569 genes desregulados en *Ccm2* KO se identificaron, en línea con resultados obtenidos anteriormente, vías de señalización relacionadas con metabolismo energético, función endotelial y señalización celular mediante MAPKs (**Figura R33C**).

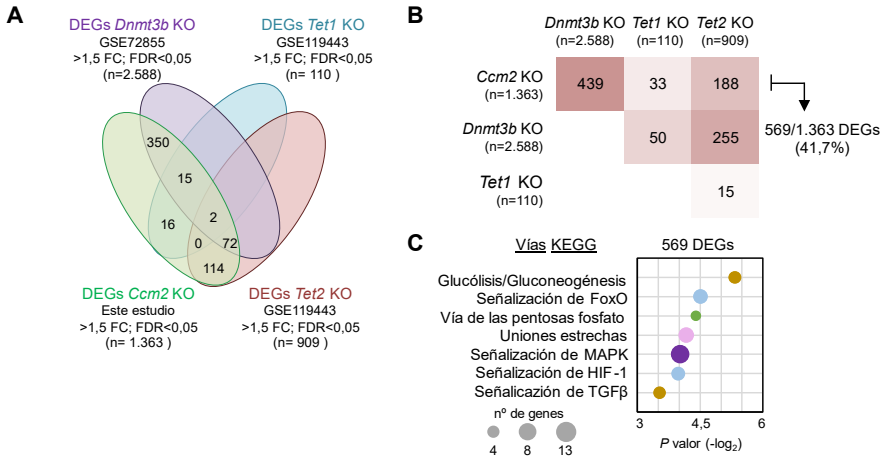


Figura R33. DNMT3B, TET1 y TET2 participan en el control de la regulación transcripcional de genes desregulados en ausencia de CCM2. A) Representación mediante diagrama de Venn de los genes diferencialmente expresados (DEGs) en ESCs KO para *Ccm2*, *Dnmt3b*, *Tet1* o *Tet2* en condiciones estándar de cultivo. Se emplearon los datos de RNA-seq generados en esta tesis para las *Ccm2* KO ESCs y los procedentes de los estudios de Neri y colaboradores (GSE72855)[294] para *Dnmt3b* KO ESCs y Reimer y colaboradores para las ESCs KO de *Tet1* y *Tet2* (GSE119443)[295]. **B)** Representación mediante matriz de similitud de los genes diferencialmente expresados (FC>1,5 y FDR<0,05) compartidos entre los pares de ESCs indicadas. Se muestra el número de genes desregulados comunes en cada comparación. **C)** Análisis de las vías de señalización KEGG empleando 569 genes desregulados indicados en **B**. Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de genes involucrados en cada vía. Los términos representados se seleccionaron entre los más enriquecidos.

En conjunto, mis resultados muestran una reducción de los niveles globales de metilación del ADN en ausencia de CCM2. A este respecto, este estado de hipometilación genómico de las ESCs KO para *Ccm2* correlaciona con la disminución de DNMT3B/TET1 y el aumento de TET2 detectados en estas células y con las funciones

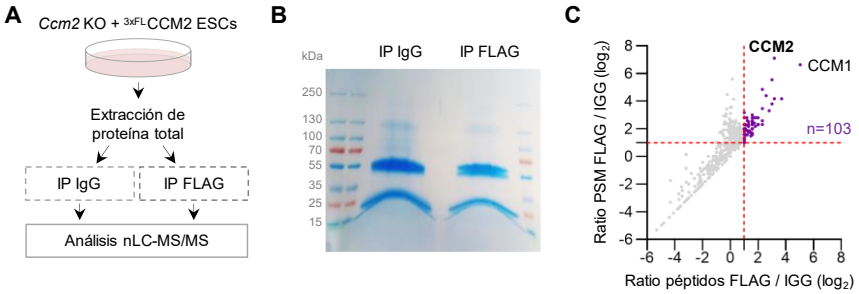
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

previamente descritas en la literatura de estos reguladores epigenéticos con respecto a la regulación de pluripotencia y de los estados alternativos *naïve* y *primed*. Además, los resultados obtenidos apoyan que tanto DNMT3B como TET1 y TET2 son necesarios para la regulación transcripcional adecuada de una importante fracción de genes cuya expresión se encuentra significativamente alterada tras la pérdida de CCM2 en ESCs.

5.2.9 Identificación de los interactores de CCM2

Para continuar explorando los mecanismos moleculares por los que CCM2 participa en el control de los estados alternativos de pluripotencia, decidí caracterizar las proteínas capaces de unirse a CCM2 en ESCs, tanto de manera directa como indirecta. Para ello, utilicé una estrategia de purificación por afinidad y espectrometría de masas (AP-MS). Brevemente, a partir de proteína total obtenida de las ESCs KO para *Ccm2* rescatadas con ^{3xFL}CCM2 procedí a inmunoprecipitar (IP) las proteínas unidas a esta proteína mediante el uso de un anticuerpo contra FLAG y, en paralelo con los mismos extractos proteicos, utilicé el anticuerpo isotipo IgG como control negativo (**Figura R34A**). Seguidamente, los complejos proteicos inmunoprecipitados mediante FLAG e IgG se separaron por tamaño para posteriormente ser procesados y analizados mediante espectrometría de masas en tándem por cromatografía de nanolíquidos (nLC-MS/MS) (**Figuras R34A** y **R34B**). Con el fin de aumentar la confianza en la identificación de las proteínas que interactúan con CCM2, apliqué los siguientes criterios: i) el número de péptidos y el número total de secuencias peptídicas (PSM, del inglés *peptide spectrum matches*) identificados en la muestra FLAG-IP debe ser superior a 1 y ii) la ratio entre el número de péptidos distintos y la ratio PSM debe ser igual o mayor a 2 entre las muestras inmunoprecipitadas con FLAG y las inmunoprecipitadas con IgG. A través de esta estrategia se identificaron 102 proteínas con elevada confianza de ser verdaderos interactores de CCM2 (**Figura R34C**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



D

Ranquin	Proteína	Ratio pépt.	Ratio PSM
1	CCM1	33	99
24	FN1	3,33	5,33
29	SETDB1	4	5
45	MED1	3	4
58	KDM3A	2,86	3,28
60	SPTBN1	2,08	3,04
70	TJP1	2	3

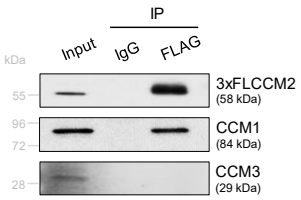
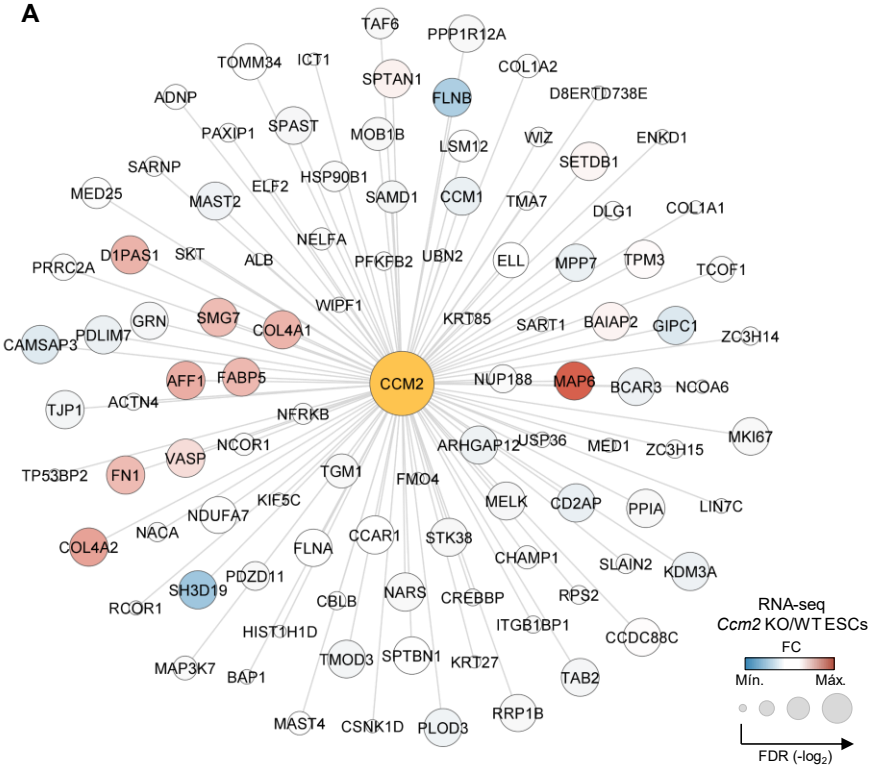


Figura R34. Identificación del interactoma de CCM2 en ESCs. A) Representación esquemática de la estrategia seguida para identificar las proteínas que interaccionan con CCM2. Se extrajo proteína total de ESCs *Ccm2* KO +^{3xFL}CCM2, se llevó a cabo una inmunoprecipitación (IP) de FLAG o IgG como control negativo y se realizó un análisis de espectrometría de masas en tándem por cromatografía de nanolíquidos (nLC-MS/MS). **B)** Imagen de la separación de las IPs indicadas en un gel Tris-Glicina de acrilamida/bisacrilamida con gradiente del 4% al 20%. **C)** Representación mediante gráfico de dispersión de las proteínas identificadas por espectrometría de masas. Se muestran CCM2 y otras 102 proteínas que superan los umbrales establecidos y son consideradas interactores, entre las que se encuentra CCM1. **D)** Proteínas interactoras con las posiciones que ocupan en el ranquin y sus ratios de péptidos y PSM FLAG/IgG. **E)** Análisis mediante IP y western blot con los anticuerpos indicados a partir de extractos proteicos totales obtenidos de la línea rescatada para CCM2 en ESCs.

Hay que destacar que el principal interactor ranqueado de CCM2 resultó ser CCM1 (**Figura R34D**), cuya interacción fue validada por IP y western blot (WB) (**Figura R34E**). Por el contrario, tanto la aproximación por AP-MS como por IP-WB no reveló la existencia de interacción entre CCM2 y CCM3 (**Figuras R34C** y **R34E**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A



B

Ranquin	Interactor	FC	FDR
21	MAP6	1,57	1,06E-14
66	COL4A2	1,00	2,04E-04
76	AFF1	0,90	6,17E-28
7	D1PAS1	0,85	1,69E-11
41	COL4A1	0,84	1,44E-06
91	FABP5	0,81	1,88E-06
24	FN1	0,78	1,47E-12
39	SMG7	0,78	6,15E-37
9	SH3D19	-0,87	4,41E-28
59	FLNB	-0,76	2,35E-38

Figura R35. Análisis del efecto de la pérdida de CCM2 sobre la expresión de los genes que codifican a sus interactores. A) Representación de la interacción con CCM2 de las 102 proteínas identificadas por AP-MS. Se integran, mediante el software *Cytoscape*, los datos de expresión diferencial entre ESCs KO para *Ccm2* y ESCs WT obtenidos en este estudio por RNA-seq. El valor de cambio (FC) determina la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

intensidad del color de las burbujas (rojo equivale a activación; azul equivale a represión). El valor de significancia estadística (-$\log_2$ FDR) determina el tamaño de la burbuja (mayor tamaño equivale a mayor significancia). **B)** Tabla con las proteínas que ven afectada su expresión génica (>1,5 FC y FDR<0,05) en ausencia de CCM2 (10/102; 9,8%) con las posiciones que ocupan en el ranquin establecido de interactores de CCM2. Se especifica el FC y FDR relativo a su alteración transcripcional.

Luego, para evaluar el comportamiento de los genes que codifican las proteínas que interaccionan con CCM2 y determinar si su expresión se encuentra alterada en su ausencia, empleé la herramienta *Cytoscape*. Los datos obtenidos mostraron que sólo 10 de los genes que codifican para las 102 proteínas que interaccionan con CCM2 (9,8%) encontraban su expresión alterada (FC>1,5 y FDR<0,05) en las ESCs KO para *Ccm2* (**Figuras R35A** y **R35B**).

Para explorar las funciones en las que intervienen los interactores de CCM2 realicé, en primer lugar, un análisis de ontología de fenotipos de mamífero (MPO, del inglés *mammalian phenotype ontology*) a través de STRING (https://string-db.org/). De manera acorde con las funciones conocidas de CCM2, se observa que la alteración de sus interactores se relaciona con términos de letalidad embrionaria, angiogénesis anormal, desarrollo vascular anormal y otras alteraciones relacionadas (**Figura R36A**). Después, a través de un análisis de ontología génica determiné los procesos biológicos relacionados con los genes que codifican los interactores de CCM2. En línea con los resultados anteriores, se observaron procesos afectados significativamente relacionados con la angiogénesis y la formación de vasos sanguíneos, así como otros relacionados con la diferenciación celular y alteraciones epigenéticas que afectan al estado de metilación de histonas (**Figura R36B**).

En conjunto, los datos obtenidos revelan las proteínas que interaccionan con CCM2 en ESCs y que potencialmente podrían intervenir de manera directa en la modulación de sus funciones

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

descritas a lo largo de esta tesis en el contexto del mantenimiento del estado celular pluripotente.

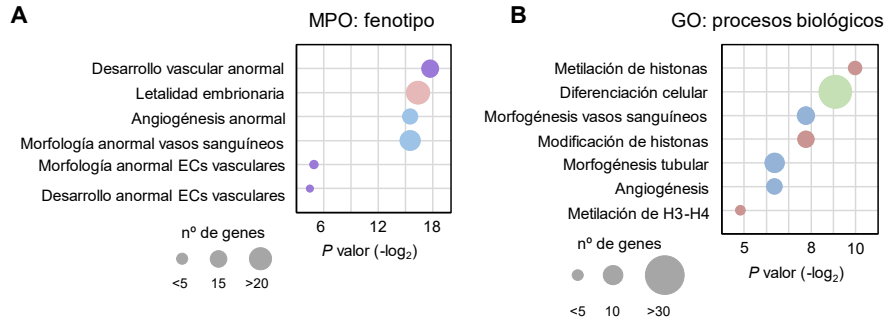


Figura R36. Análisis de los procesos y fenotipos en los que intervienen los interactores de CCM2. A) Análisis de ontología de fenotipos de mamíferos (MPO) empleando la lista que contiene a CCM2 y sus 102 interactores. Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de interactores involucrados en cada fenotipo. **B)** Análisis de ontología génica (GO) de procesos biológicos igual que en **A**. En **A** y **B** los términos representados se seleccionaron entre los más enriquecidos.

5.2.10 Predicción de interactores de CCM2 envueltos en la regulación de pluripotencia *naïve*

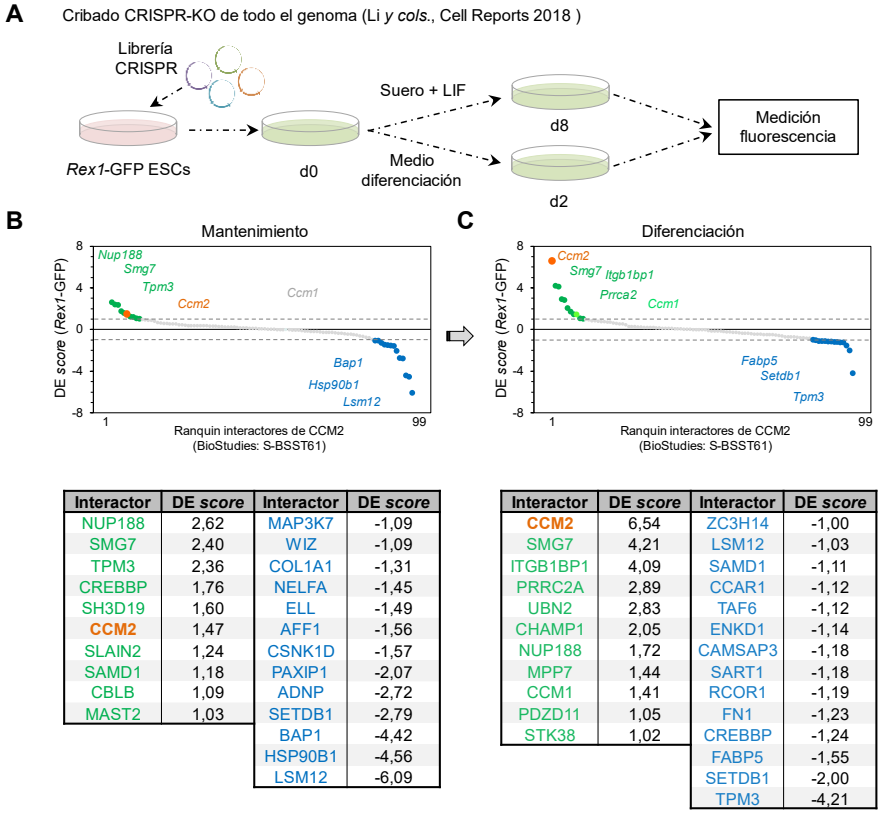
Dado que la pérdida de CCM2 en ESCs promueve un estado pluripotente con características *naïve*, lo siguiente que me propuse fue explorar cuáles de los interactores de CCM2 podrían ejercer también un papel similar restringiendo este estado alternativo de pluripotencia. Para este fin, a partir de los genes que codifican a *Ccm2* y sus interactores, realicé un análisis *in silico* usando los datos de un cribado de pérdida de función a escala global mediante el sistema CRISPR-Cas9 para la identificación de reguladores potenciales directos e indirectos de la expresión del gen *Rex1*, marcador de pluripotencia *naïve*[75,296], durante el mantenimiento y la salida de pluripotencia[297]. A partir de los valores de la expresión diferencial (DE *score*) de *Rex1*-GFP superiores a 1 pude identificar posibles genes candidatos cuya pérdida de función favorece (altos niveles de *Rex1*-GFP, DE>1) o

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

dificulta (bajos niveles de *Rex1*-GFP, DE>-1) la adquisición o retención del estado *naïve* en condiciones estándar de cultivo (SL) o en medio de diferenciación, respectivamente (**Figura R37A**). De manera destacada, en condiciones de mantenimiento estándar (SL) de las ESCs *Rex1*-GFP, se observó que la pérdida de función de CCM2 incrementaba la actividad reportera de GFP (DE=1,47), lo que es consistente con los datos obtenidos a lo largo de este trabajo indicando que las ESCs KO para *Ccm2* promueven un estado pluripotente similar al *naïve* (**Figura R37B**). Además, se identificaron otros 9 interactores (NUP188, SMG7, TPM3, CREBBP, SH3D19, SLAIN2, SAMD1, CBLB y MAST2) que, de manera similar a CCM2, favorecían el aumento de los niveles de *Rex1*-GFP tras su eliminación (DE>1), mientras que también se identificaron 13 interactores con el efecto opuesto a CCM2, facilitando tras su pérdida la disminución de los niveles de GFP (DE>-1) (**Figura R37B**). Por otro lado, se observó que, tras forzar la diferenciación, la pérdida de CCM2 retenía pluripotencia *naïve* (DE>6,54) (**Figura R37C**), lo que también es consistente con mis resultados previos. Interesantemente, bajo diferenciación se observó que la pérdida de CCM1 favorecía el aumento de la actividad del reportero GFP (DE=1,41) (**Figura R37C**), lo que sugiere que esta proteína CCM puede ser relevante durante el proceso de diferenciación. Con respecto a la pérdida de función de otros interactores de CCM2, se identificaron 9 (SMG7, ITGB1BP1, PRRC2A, UBN2, CHAMP1, NUP188, MPP7, PDZD11 y STK38) que también favorecían la retención de pluripotencia *naïve* en condiciones de diferenciación (DE>1) y 14 que tenían el efecto contrario (DE>-1) (**Figura R36C**).

En conjunto, los datos obtenidos revelan varios interactores de CCM2 que pueden ser reguladores potenciales directos o indirectos de la expresión del gen *Rex1* tanto en condiciones relacionadas con el mantenimiento como durante la salida de pluripotencia *naïve*. Serán necesarios futuros estudios para confirmar estos hallazgos y determinar qué interactores cooperan con CCM2 para regular el mantenimiento y salida de pluripotencia.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



Interactor	DE *score*	Interactor	DE *score*
NUP188	2,62	MAP3K7	-1,09
SMG7	2,40	WIZ	-1,09
TPM3	2,36	COL1A1	-1,31
CREBBP	1,76	NELFA	-1,45
SH3D19	1,60	ELL	-1,49
CCM2	1,47	AFF1	-1,56
SLAIN2	1,24	CSNK1D	-1,57
SAMD1	1,18	PAXIP1	-2,07
CBLB	1,09	ADNP	-2,72
MAST2	1,03	SETDB1	-2,79
		BAP1	-4,42
		HSP90B1	-4,56
		LSM12	-6,09

Interactor	DE *score*	Interactor	DE *score*
CCM2	6,54	ZC3H14	-1,00
SMG7	4,21	LSM12	-1,03
ITGB1BP1	4,09	SAMD1	-1,11
PRRC2A	2,89	CCAR1	-1,12
UBN2	2,83	TAF6	-1,12
CHAMP1	2,05	ENKD1	-1,14
NUP188	1,72	CAMSAP3	-1,18
MPP7	1,44	SART1	-1,18
CCM1	1,41	RCOR1	-1,19
PDZD11	1,05	FN1	-1,23
STK38	1,02	CREBBP	-1,24
		FABP5	-1,55
		SETDB1	-2,00
		TPM3	-4,21

Figura R37. Cribado *in silico* de CCM2 e interactores en la regulación del gen *naïve Rex1*. A) Representación esquemática de la estrategia seguida en el estudio de Li y colaboradores para analizar el mantenimiento y la pérdida de pluripotencia *naïve*[297]. Se realizó un cribado empleando una librería CRISPR-KO en ESCs que expresan el marcador fluorescente GFP bajo el control del marcador *naïve Rex1*. Analizando la intensidad de fluorescencia se calculó la expresión diferencial de GFP (DE score) tras la eliminación de cada uno de los genes diana de la librería. Se detectaron genes que afectan: i) a la adquisición y mantenimiento del estado *naïve* tras 8 días en condiciones estándar de cultivo (SL) y ii) al proceso de salida del estado pluripotente tras 2 días en condiciones de diferenciación. **B)** (Arriba) Representación gráfica de los DE score ordenados tras la pérdida de función de CCM2 y cada uno de sus interactores tras mantener las ESCs *Rex1*-GFP 8 días en SL. Se resaltan junto con CCM2 y CCM1 los tres principales interactores cuya pérdida favorece el mantenimiento del estado *naïve* (verde) o lo desfavorece (azul). (Abajo) Tabla con las proteínas interactoras de CCM2 con DE score >1. **C)** (Arriba) Representación gráfica de los DE score ordenados tras la eliminación de CCM2 y cada uno de sus interactores tras mantener las ESCs *Rex1*-GFP 2 días en condiciones de diferenciación. Se resaltan junto con CCM2 y CCM1 los tres principales interactores

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cuya pérdida resulta en la retención del estado *naïve* (verde) o que favorecen la salida de pluripotencia (azul). (Abajo) Tabla con las proteínas interactoras de CCM2 con DE score >1.

5.2.11 CCM2 vincula múltiples vías epigenéticas a la red de pluripotencia

La regulación del control de la expresión génica a nivel epigenético resulta esencial en el control de los programas moleculares que gobiernan el mantenimiento y el cambio de la identidad celular pluripotente[52,69,78,79]. A este respecto, mis resultados anteriores indican que CCM2 participa de manera indirecta en el control transcripcional en ESCs a través de la regulación epigenética mediada por el complejo represor PCR2 (**Figuras R26-R28**) y los moduladores del estado de metilación en cromatina DNMTs/TETs (**Figuras R30-R33**). Teniendo esto en cuenta, me propuse explorar si en el papel de la regulación de la expresión génica gobernada por CCM2 en ESCs intervienen mecanismos epigenéticos intrínsecos a su red de interactores proteicos. A partir de una lista actualizada que contiene 914 proteínas de las que se ha descrito que participan en la regulación epigenética (https://epifactors.autosome.org/), pude detectar 14 reguladores epigenéticos entre los 102 interactores de CCM2 (13,75%) (**Figura R38A**). De manera destacada, se observó que CCM2, a través de estos 14 interactores, está conectado con múltiples componentes reguladores de vías epigenéticas con documentada significancia funcional en pluripotencia (**Figura R38B**). En concreto, se identificaron componentes de complejos epigenéticos involucrados en procesos de remodelación de la cromatina (ADNP, NFRKB y RCOR1), de acetilación (CREBBP, MAP3K7 y TAF6) y desacetilacion (NCOR1 y RCOR1) de histonas, de ubiquitinación de histonas (BAP1 y USP36) y de metilación de histonas (KDM3A, NCOA6, PAXIP1, RCOR1, SETDB1 y TAF6) (**Figura R38B**). Para explorar el efecto de la asociación física de CCM2 con estos factores decidí examinar si algunas de sus marcas epigenéticas asociadas a histonas podrían verse afectadas a nivel global en ausencia de CCM2. Interesantemente, se

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

observó que al contrario de H3K27me3, mediada por PRC2, los niveles globales en cromatina para H3K27ac y H3K9ac se encontraban ligeramente disminuidos en las ESCs KO para *Ccm2* comparado con las ESCs WT en condiciones estándar de cultivo (SL) (**Figura R38C**).

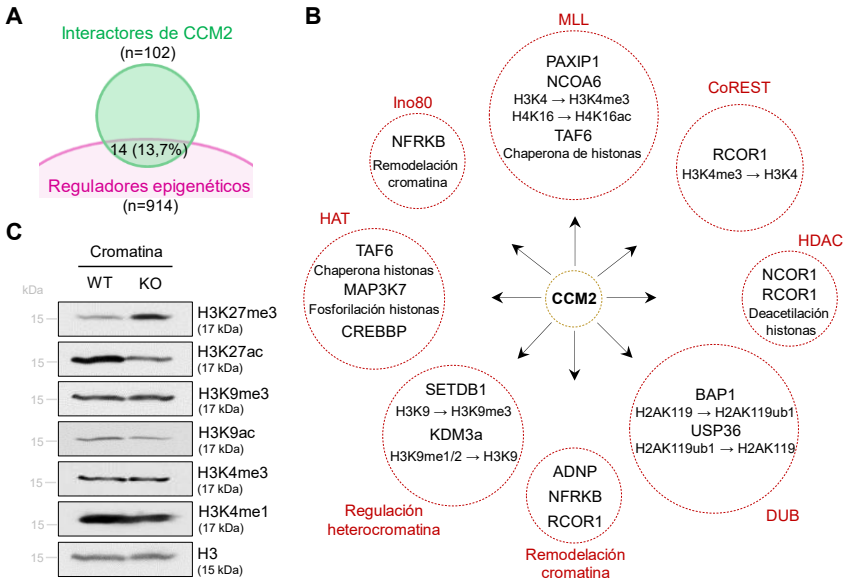


Figura R38. CCM2 interacciona con varios complejos que regulan la cromatina. A) Representación mediante diagrama de Venn de los interactores de CCM2 y una lista que contiene 914 componentes de complejos epigenéticos. **B)** Representación de la conexión de CCM2 con múltiples vías de regulación epigenética a través de sus interactores. Se representan los principales complejos proteicos (círculos de puntos rojos) y se muestran las modificaciones de histonas asociadas a los interactores. **C)** Western blot para detectar los niveles globales de las marcas de histonas H3K27me3, H3K27ac, H3K9me3, H3K9ac, H3K4me3 y H3K4me1 en la cromatina de ESCs WT o KO para *Ccm2* en condiciones estándar de cultivo. H3 se empleó como control de carga.

Finalmente, tras apreciar una asociación física funcional entre CCM2 y complejos epigenéticos, decidí analizar si los 1.363 genes diferencialmente expresados (FC>1,5 y FDR<0,05) en ESCs con pérdida de CCM2 podrían tener afectadas sus regiones promotoras y potenciadoras (*enhancers* en inglés). Con este fin, realicé una inmunoprecipitación de la cromatina unida a las marcas H3K4me1,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

H3K4me3, H3K27me3 y H3K27ac seguida de secuenciación de ADN (ChIP-seq) en ESCs WT y KO para *Ccm2*. Desafortunadamente, durante el proceso de generación de las librerías de secuenciación se produjo una sobreamplificación de los productos de PRC para H3K27me3 y H3K27ac, que no ha permitido tener a tiempo datos de secuenciación fiables para estas dos marcas de histonas en el transcurso de la elaboración de este manuscrito de tesis. Sin embargo, a partir de los datos de calidad de ChIP-seq obtenidos para H3K4me3 (denota regiones promotoras) y H3K4me1 (denota regiones potenciadoras) procedí a analizar cómo se encontraban estas marcas de histonas en genes cuya expresión se encuentra afectada en las ESCs como consecuencia de la ausencia de CCM2.

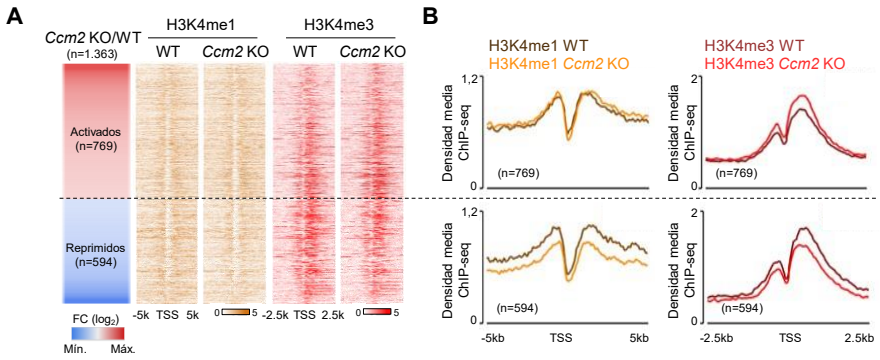


Figura R39. La ausencia de CCM2 afecta a regiones potenciadoras y promotoras en genes desregulados. A) Representación mediante mapa de calor de la densidad de unión de H3K4me1 y H3K4me3 a las regiones potenciadoras y promotoras de los 1.363 genes desregulados en ausencia de CCM2 en ESCs WT y KO bajo condiciones estándar de cultivo (SL). Se estableció una ventana de 10 kb alrededor del TSS para las regiones potenciadoras y otra de 5 kb para las promotoras. Los genes desregulados se muestran ordenados por valor de cambio en su expresión (FC) de mayor a menor. **B)** Representación mediante histograma de la densidad media de unión de H3K4me1 y H3K4me3 en los genes ordenados igual que en **A**. Se emplearon los datos de ChIP-seq generados en esta tesis.

Para caracterizar las regiones potenciadoras analicé la abundancia de la marca H3K4me1 en una ventana de 10 kb alrededor del sitio de inicio de la transcripción (TSS), mientras que para las

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

regiones promotoras analicé la marca H3K4me3 en una ventana de 5 kb centrada en el TSS de los genes desregulados (**Figuras R39A** y **R39B**). De manera importante, se pudo apreciar que en las ESCs KO para *Ccm2*: i) la activación de los 769 genes correlaciona principalmente con el aumento en la deposición de H3K4me3 (parte superior, **Figuras R39A** y **R39B**) y ii) la represión de los 594 genes correlaciona con una clara disminución en los niveles tanto de H3K4me3 como de H3K4me1 (parte inferior, **Figuras R39A** y **R39B**).

En conjunto, los datos obtenidos establecen una sólida base bioquímica para continuar explorando cómo CCM2, a través de su vínculo con múltiples vías epigenéticas, es capaz de contribuir a modular el paisaje transcripcional de las ESCs.

5.3 DETERMINAR LA FUNCIÓN CCM EN DIFERENCIACIÓN Y ESPECIFICACIÓN ENDOTELIAL

5.3.1 La función CCM participa en la regulación de programas transcripcionales de diferenciación celular

Además de la capacidad de autorrenovación, la segunda característica que define a las células pluripotentes es su habilidad de poder dar lugar a cualquiera de los tipos celulares que conforman un individuo adulto. Mientras que la pérdida de función individual de cada gen *Ccm* resulta en letalidad embrionaria alrededor de E8.5 y E10 (para más detalle ver sección 2.4.2), periodo en el que se aprecian diferencias morfológicas y fallos angiogénicos y de circulación[178,188,217,219], el papel concreto de cada proteína CCM durante la salida de pluripotencia en desarrollo embrionario aún no se ha investigado. Teniendo en cuenta que las ESCs representan un excelente modelo en cultivo para investigar los procesos moleculares que intervienen en las etapas iniciales de especificación celular durante el desarrollo embrionario temprano, me propuse estudiar el efecto de la pérdida de cada uno de los genes *Ccm* durante los procesos de diferenciación en cultivo celular. En primer lugar, analicé las dinámicas de expresión de *Ccm1*, *Ccm2* y

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Ccm3 durante el proceso de diferenciación a partir de ESCs murinas utilizando datos de expresión previamente publicados. En concreto, durante la diferenciación de ESCs con ácido retinoico (RA, del inglés *retinoic acid*)[298], se vio que *Ccm1* y *Ccm2* se encontraban regulados al alza a medida que avanzaba el proceso de diferenciación, mientras que *Ccm3* vio reducidos sus niveles tras la salida de pluripotencia (**Figura R40A**). En el caso del proceso de diferenciación mediante la generación de cuerpos embrioides (EBs, del inglés *embryoid body*)[299], se pudo apreciar que: i) los niveles de *Ccm1* mostraban un pico de expresión entre las 24 y 48 horas tras iniciarse la diferenciación, ii) los niveles de *Ccm2* se elevaban desde el inicio hasta el final del proceso de diferenciación y iii) la expresión de *Ccm3* se vio reducida a medida que avanzaba el proceso (**Figura 40B**). Estos datos sugieren que cada gen *Ccm* puede desempeñar funciones específicas a lo largo del proceso de diferenciación que se emplee. Para analizar si los genes *Ccm* podrían estar implicados en el control transcripcional que coordina la salida de pluripotencia realicé un análisis mediante GSEA empleando una lista de genes involucrados en la diferenciación de ESCs (GO:0048863). De manera destacada, se observó que estos genes que participan en el proceso de diferenciación celular se encontraban enriquecidos en las ESCs WT en comparación con los KO para *Ccm1* (NES=1,98 y FDR=0,0) y *Ccm3* (NES=2,31 y FDR=0,0), mientras que se mostraban enriquecidos en el KO para *Ccm2* (NES=-1,53 y FDR=0,0) (**Figura R40C**). Estos resultados apoyan que la pérdida de función de cada gen *Ccm* en ESCs podría alterar el correcto proceso de toma de decisiones celulares durante la salida de pluripotencia. Además, revela que CCM2 podría ejercer funciones específicas y distinguibles de CCM1 y CCM3 para la correcta diferenciación celular a partir de ESCs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

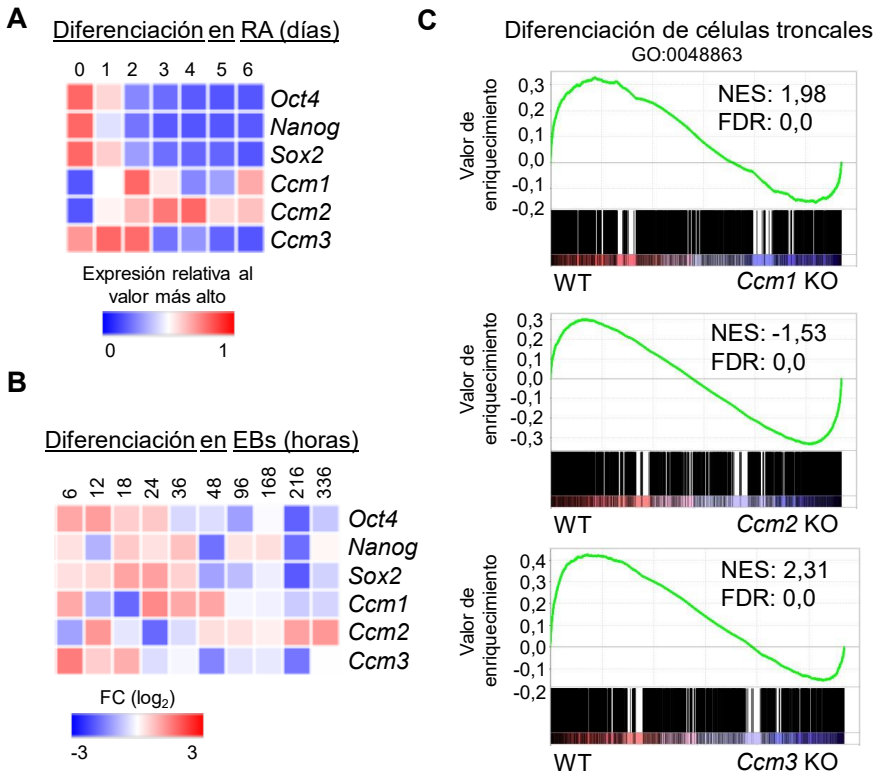


Figura R40. La pérdida de función CCM altera programas transcripcionales de diferenciación celular en ESCs. A) Representación mediante mapa de calor de la expresión relativa de los genes indicados durante 6 días de diferenciación de ESCs en presencia de ácido retinoico (RA). Se reanalizaron los datos del estudio de Ivanova y colaboradores (GSE4679)[298]. **B)** Representación mediante mapa de calor de los cambios en la expresión relativa de los genes indicados durante 336 horas de diferenciación de ESCs hacia cuerpos embrioides (EBs). Se reanalizaron los datos del estudio de Hailesellasse Sene y colaboradores (GSE3749)[299]. **C)** Análisis GSEA de los datos de RNA-seq de ESCs WT y KO para *Ccm1* (arriba), *Ccm2* (centro) y *Ccm3* (abajo) generados en este estudio en condiciones estándar de cultivo contra una lista de genes involucrados en la diferenciación de células troncales (GO:0048863). Se muestran los valores de NES y FDR.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.3.2 Las ESCs deficientes en CCMs muestran una diferenciación anormal hacia cuerpos embrioides

Dado el papel potencial que las proteínas CCM pueden ejercer sobre los procesos de diferenciación celular, decidí analizar las consecuencias de su ausencia en el proceso de formación de cuerpos embrioides (EBs). Este ensayo de diferenciación mimetiza en cultivo los procesos tempranos del desarrollo embrionario y la especificación inicial hacia las tres capas germinales (ectodermo, mesodermo y endodermo)[251]. Con este fin, ESCs WT y KO para *Ccm1*, *Ccm2* o *Ccm3* se cultivaron en suspensión en medio estándar sin LIF durante 9 días, lo que resulta suficiente para generar los agregados multicelulares en 3 dimensiones (EBs) (**Figura R41A**). En línea con los resultados *in silico* (**Figura R40B**), tras 9 días (216 horas) de diferenciación: i) los niveles de expresión *Ccm2* se encontraban regulados al alza de manera similar al marcador de ectodermo *Pax3* y al marcador de endodermo *Gata6* y ii) los niveles de *Ccm3* estaban disminuidos al igual que los del marcador de pluripotencia *Nanog* (**Figura R41B**). Luego, mediante análisis morfológico por microscopía se observó que los agregados en 3D generados a partir de las ESCs KO no presentaban las características evaginaciones (*cysts*) de los EBs formados a partir de ESCs WT tras 9 días de diferenciación (**Figura R41C**). Además, la ausencia de cada proteína CCM afectó a la longitud de los EBs generados, siendo estos más pequeños que los WT tras 9 días (**Figura R41D**). En línea con estos datos, que indican que la pérdida de cada gen *Ccm* conlleva EBs más pequeños y menos desarrollados, el análisis mediante qPCR de la expresión génica de distintos marcadores de pluripotencia y de cada una de las tres capas germinales confirmó aberraciones transcripcionales en los EBs KO comparados con los EBs WT (**Figura R41E**). En concreto, se observó que cada gen *Ccm* es necesario para: i) el correcto silenciamiento del marcador de pluripotencia *Nanog*, en particular en el caso de *Ccm2* KO, y ii) la activación de los marcadores de ectodermo (*Sox1*, *Pax3* y *Nestin*), en donde los defectos fueron más evidentes en *Ccm1* KO y *Ccm3* KO, y iii) los niveles correctos de marcadores de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

mesodermo (*Cxcl12*, *Mixl1*, *Eomes*, *T* y *Mest*) y endodermo (*Gata4*, *Gata6*, *Sox17* y *Afp*), en donde se aprecia de manera especial que la ausencia de CCM2 favoreció la regulación al alza de genes de estas dos capas germinales (**Figura R41E**). Estos datos muestran que la función CCM es necesaria durante diferenciación celular y que las ESCs KO para cada gen *Ccm* pueden estar sesgadas hacia la diferenciación de mesodermo y endodermo durante la formación de EBs.

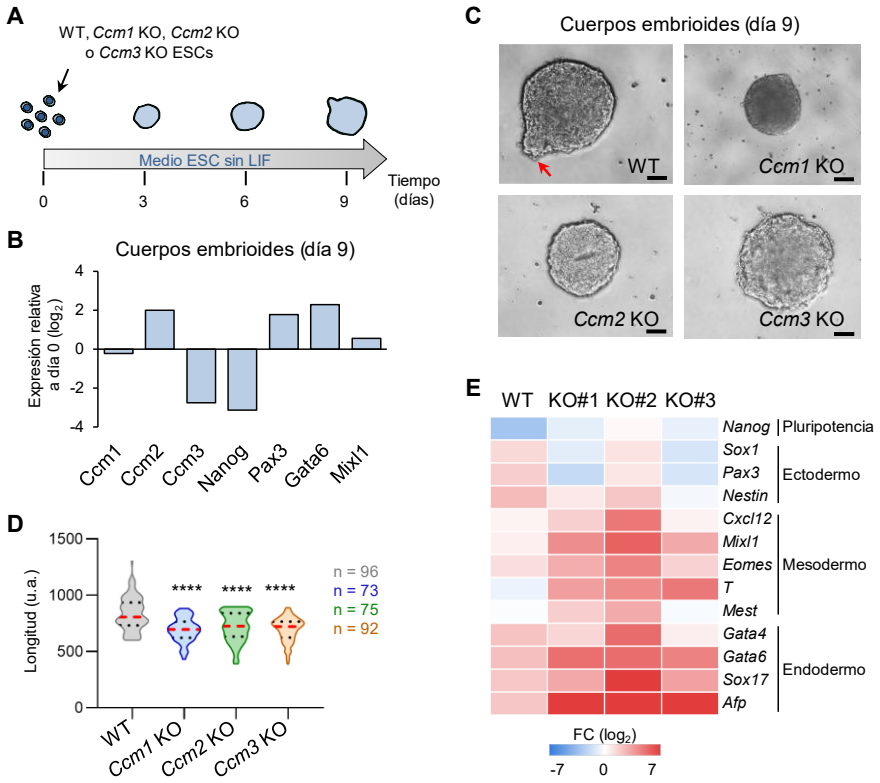


Figura R41. Las proteínas CCM son necesarias durante la formación de EBs. A) Representación gráfica de la estrategia seguida para diferenciar células pluripotentes hacia cuerpos embrioides (EBs). Se sembraron ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en medio sin LIF en placas de baja adherencia, y se dejaron crecer durante 9 días para la formación de agregados con evaginaciones. **B)** Representación mediante gráfico de barras de la expresión relativa de los genes *Ccm*, el marcador de pluripotencia *Nanog*, el de ectodermo *Pax3*, el de endodermo *Gata6* y el de mesodermo *Mixl1* a día 9 de formación de EBs. **C)** Imágenes representativas

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de los EBs generados a día 9 a partir de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3*. Se señala con una flecha la evaginación generada en condiciones WT, estructura ausente en condiciones KO. Las barras representan 50 µm. **D)** Representación mediante gráfico de violín de la longitud (en unidades arbitrarias) de los EBs generados a día 9 a partir de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3*. Se indica el número total de EBs cuantificados. Se representa la mediana para cada conjunto de datos con una línea punteada roja, y los cuartiles con líneas negras. Se aplicó la prueba ANOVA de un factor para el análisis de la varianza y la prueba de comparaciones múltiples de Dunnett para el cálculo estadístico entre cada KO y WT. ****p<0,0001. **E)** Representación mediante mapa de calor del cambio en la expresión relativa de marcadores pluripotentes y específicos de capa germinal entre los días 9 y 0 en células WT y KO para *Ccm1*, *Ccm2* o *Ccm3*.

5.3.3 La eliminación de las CCM compromete la gastrulación en cultivo

Con el fin de confirmar los resultados anteriores en los que la pérdida de CCM1, CCM2 o CCM3 afecta a la correcta especificación de las capas germinales embrionarias en cultivo, decidí realizar un ensaño de formación de gastruloides. Estas estructuras consisten en agregados celulares en 3D que sufren procesos similares a la gastrulación y la elongación *in vivo*, mimetizando la expresión génica espacial y temporal del embrión[252]. Con este fin, ESCs WT y KO para *Ccm1*, *Ccm2* o *Ccm3* se cultivaron en suspensión en pocillos con fondo en U en medio sin suero (N2B27) durante 120 horas. Entre las 48 y 72 horas se añadió un inhibidor de GSK3β (CHIR99021) (**Figura R42A**). Transcurridas 120 horas, las células han pasado por procesos de agrupación, elongación y polarización, para dar lugar al inicio de la curvatura de estas estructuras. De manera destacada, el análisis morfológico por microscopía revela que la ausencia de cada gen *Ccm* afectó a la correcta formación de estructuras gastruloides en cultivo (**Figura R42B**). En particular, la ausencia de CCM2 provocó defectos en la elongación y polarización causando una reducción significativa en la longitud de estas estructuras en 3D (**Figura R42C**). A continuación, determiné la eficacia que presenta cada línea celular para formar gastruloides a través de su porcentaje de formación. Así, aun siendo capaces de generar estructuras en 3D, las ESCs KO para *Ccm1*, *Ccm2*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

o *Ccm3* tienen una eficiencia significativamente reducida para formar gastruloides en cultivo comparadas con las ESCs WT (**Figura R42D**).

En conjunto, mis resultados muestran que las CCMs son importantes durante el proceso de gastrulación en cultivo, lo que correlacionaría con defectos que pudiesen causar letalidad embrionaria observada en las etapas posteriores a gastrulación en los ratones KO para *Ccm1*, *Ccm2* y *Ccm3*. Serán necesarios futuros estudios para confirmar estos efectos potenciales de la función CCM durante las etapas de post-implantación embrionaria *in vivo*.

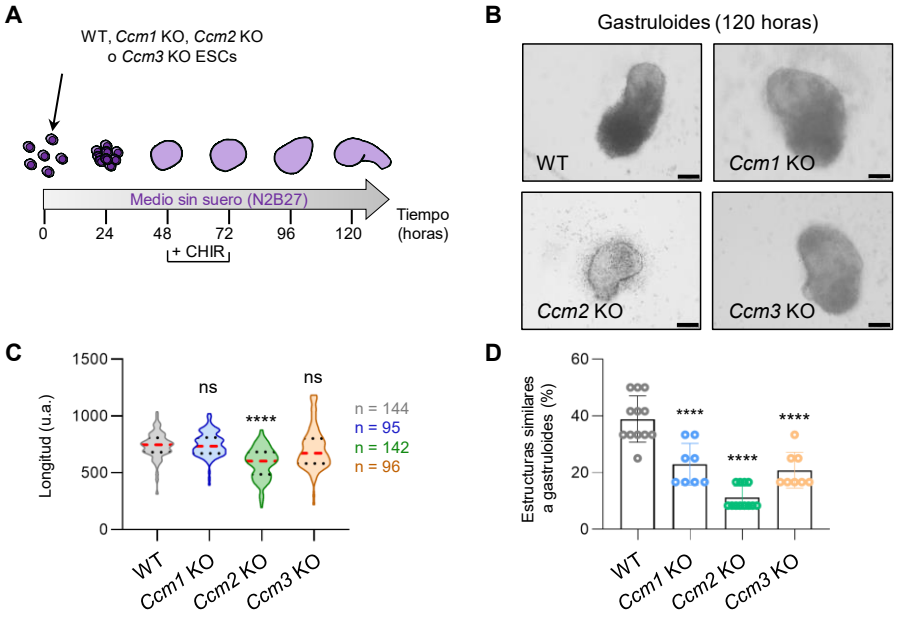


Figura R42. Las CCM son necesarias para una correcta formación de gastruloides. **A)** Representación gráfica de la estrategia seguida para diferenciar las ESCs hacia gastruloides. Se sembraron ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en medio sin suero (N2B27) en placas de baja adherencia con fondo en U, y se dejaron crecer durante 120 horas para permitir la formación de agregados tras los procesos de elongación, polarización y curvatura. Entre las 48 y 72 horas se añade el inhibidor de GSK3β (CHIR) al cultivo. **B)** Imágenes representativas de los gastruloides generados a las 120 horas a partir de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3*. Las barras representan 50 µm **C)** Representación mediante gráfico de violín de la longitud (en unidades arbitrarias) de las estructuras generadas a las 120 horas a partir de ESCs

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

WT o KO para *Ccm1*, *Ccm2* o *Ccm3*. Se indica el número total de estructuras cuantificadas. **D)** Representación mediante gráfica de barras del porcentaje de eficiencia en la generación de estructuras gastruloides generadas a las 120 horas a partir de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3*. En **C** se representa la mediana para cada conjunto de datos con una línea punteada roja, y los cuartiles con líneas negras. En **D** los datos se presentan como media ± s.e.m. En **C** y **D** se aplicó la prueba ANOVA de un factor para el análisis de la varianza y la prueba de comparaciones múltiples de Dunnett para el cálculo estadístico entre cada KO y WT. ****p<0,0001; ns: no significativo.

5.3.4 Optimización y validación de la diferenciación hacia células progenitoras endoteliales a partir de ESCs

Como se ha mostrado, la ausencia de las proteínas CCM altera el proceso de diferenciación en cultivo, dando lugar a aberraciones en la especificación de las capas germinales. Estas aberraciones sugieren que la falta de CCM1, CCM2 o CCM3 podría dificultar e incluso bloquear la obtención de tipos celulares determinados a partir de ESCs. Varios estudios han revelado la función de las proteínas CCM en el contexto endotelial, resultando su ausencia letal durante el desarrollo embrionario o provocando la aparición de malformaciones cavernosas cerebrales originadas en células endoteliales (ECs, del inglés *endothelial cells*) y/o progenitoras endoteliales (EPCs, del inglés *endothelial progenitor cells*) durante la vida postnatal[178,188,217,234-237]. El linaje endotelial deriva del mesodermo[159], capa germinal afectada por la ausencia de las proteínas CCM de acuerdo con mis resultados. Todo esto me impulsó a analizar las consecuencias de la pérdida de función CCM durante la especificación a linaje endotelial a partir de ESCs. Para ello, lo primero que me propuse fue establecer en el laboratorio un protocolo de diferenciación para obtener células progenitoras endoteliales (EPCs) que me permitiese estudiar el efecto de la ausencia de CCM1, CCM2 o CCM3 en ESCs sobre el proceso de especificación celular a linaje endotelial. La eficiencia y consecución de la especificación hacia EPCs, caracterizadas por una elevada expresión del gen *Flk1* (también denominado como *Kdr* o Vegfr2)[159,162,163], constituye un proceso altamente dependiente del tratamiento de la placa

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de cultivo, la densidad celular y el medio definido de diferenciación que se utiliza. Dada la dificultad presentada durante el transcurso de mi tesis para reproducir protocolos publicados de diferenciación endotelial en cultivo a partir de células pluripotentes murinas, decidí optimizar un protocolo de diferenciación a linaje progenitor endotelial tomando de base los diversos estudios relacionados que existían hasta la fecha[300-302]. En primer lugar, determiné la densidad celular óptima a sembrar de ESCs para llevar a cabo la diferenciación dando como resultado 13.000 células por cm^2 (**Figura R43A**). Luego, determiné el tratamiento de la placa de cultivo y el medio definido de diferenciación que permitiese obtener una mayor eficiencia en los niveles de activación del gen *Flk1* en las células resultantes. De manera resumida, tras descartar varios tratamientos y concentraciones entre los que se incluían gelatina, fibronectina y colágeno I entre otros, reduje las opciones más favorables a: a) 18 $\mu g/cm^2$ de colágeno IV, b) Matrigel#1, matriz cualificada para ESCs humanas, c) Matrigel#2, matriz de membrana basal y d) Geltrex, matriz de membrana basal de factor de crecimiento reducido. En cuanto a los diferentes medios de cultivo para diferenciación con mejor resultado analizados fueron: 1) medio estándar pluripotente (suero + LIF), 2) medio "*ESC to EC*" y 3) medio sin suero (N2B27) con ausencia de insulina. Las células se mantuvieron en cada una de las combinaciones de tratamiento de placa y medio indicados durante 4 días, con la excepción de que el medio N2B27 sin insulina fue suplementado a día 2 de diferenciación con CHIR99021, ACTIVINA A y FGFb (**Figura R43A**). Tras evaluar mediante PCR los niveles de expresión de *Flk1* a partir de ADN complementario obtenido de cada una de estas combinaciones, se determinó que la condición que más incrementaba los niveles de expresión de este marcador endotelial correspondía a medio N2B27 sobre colágeno IV (**Figuras R43B** y **R43C**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

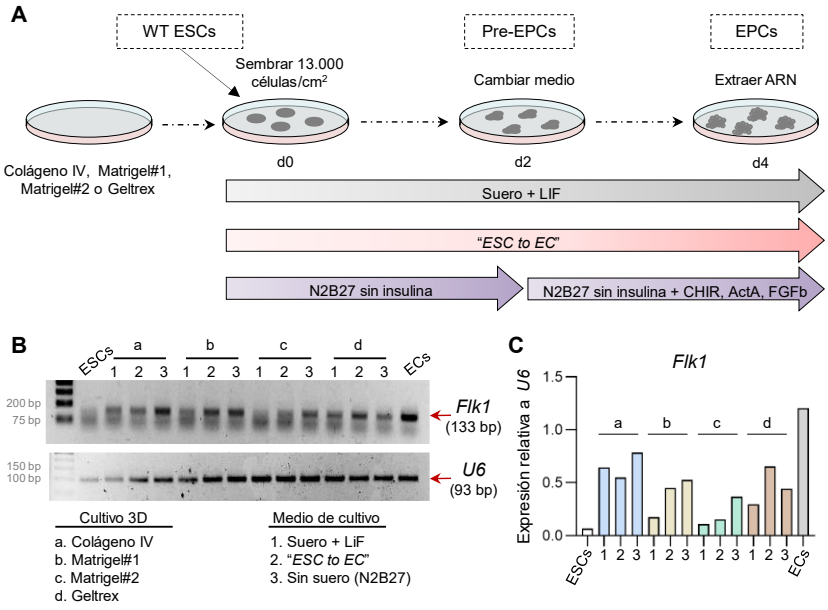


Figura R43. Determinación del protocolo óptimo de diferenciación de ESCs hacia EPCs. A) Representación gráfica de la estrategia seguida para optimizar el proceso de obtención de células progenitoras endoteliales (EPCs) a partir de ESCs de origen murino. Se sembraron 13.000 células/cm^2 de ESCs WT en placas tratadas con 18 µg/cm^2 de colágeno IV, con Matrigel#1 (matriz cualificada para hESCs), con Matrigel#2 (matriz de membrana basal) o con Geltrex (matriz de membrana de factor de crecimiento reducido). Se cultivaron las ESCs en medio estándar pluripotente (suero + LIF), medio "*ESC to EC*" y medio N2B27 sin insulina durante 4 días. El medio N2B27 se suplementó durante los últimos 2 días con CHIR (inhibidor de GSK3β), ACTIVINA A y FGFb. Tras los 4 días se extrajo ARN y se generó ADN complementario para su análisis. **B)** Análisis de la expresión de *Flk1* mediante PCR y electroforesis en gel de agarosa para todas las condiciones testadas en **A**. *U6* se empleó como control de carga. ESCs y células endoteliales maduras (ECs) se emplearon como controles negativo y positivo para la expresión de *Flk1*, respectivamente. Se indican los tamaños esperados de cada amplicón. **C)** Cuantificación mediante gráfico de barras de la expresión de *Flk1* analizada en **B** normalizada a los niveles de *U6*.

Tras seleccionar las mejores condiciones para el ensayo de especificación a EPCs (medio N2B27 sobre colágeno IV), me propuse validar las células progenitoras generadas en esta tesis. Con este fin, decidí evaluar primero si las EPCs obtenidas tras 4 días de diferenciación son capaces de generar células endoteliales. Para ello, se

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

sembraron ESCs en colágeno IV y se mantuvieron 2 días en medio N2B27 sin insulina, seguidos de 2 días en medio suplementado con CHIR99021, ACTIVINA A y FGFb. A continuación, se mantuvieron las EPCs durante dos días más en medio suplementado con FGFb, BMP4, VEGF, Br-cAMP y SB431542 (un inhibidor de los receptores tipo I ALK4/5/7 de la superfamilia TGFβ), dando lugar en el día 6 de diferenciación a células con morfología endotelial (**Figura R44A**). En segundo lugar, me propuse determinar una firma de expresión génica para definir transcripcionalmente a las EPCs. Para este fin, tomé ventaja de datos transcriptómicos disponibles mediante RNA-seq de EPCs[303] y de ECs maduras procedentes de diversos tejidos[304-306], junto con mis datos transcriptómicos de ESCs WT (**Figura R44A**). Luego, determiné los genes diferencialmente expresados (FC>1,5 y FDR<0,05) entre EPCs y ESCs y los diferencialmente expresados entre EPCs y ECs. De esta manera, pude identificar 986 genes que resultan suficientes para definir de manera robusta los tipos celulares de ESCs, EPCs y ECs (**Figura R44B**). Un análisis más en profundidad de la expresión de estos 986 genes mediante mapa de calor acoplado a un agrupamiento jerárquico reveló la existencia de patrones de expresión dinámicos entre los tres tipos celulares que comprenden dos grupos principales y cuatro clases (**Figura R44C**). El grupo 1 contiene 586 genes que se encuentran regulados al alza o a la baja en EPCs, mientras que los 400 genes del grupo 2 presentan patrones de expresión elevados en ESCs o en ECs. Dentro del grupo 1 se encuentra: clase 1, compuesta por 313 genes altamente expresados en EPCs y clase 2, que contiene 273 genes cuya expresión se reduce significativamente en este tipo progenitor celular. El grupo 2 comprende otras dos clases de genes con patrones específicos de expresión: clase 3, con 112 genes cuya expresión se encuentra significativamente elevada en las ESCs y clase 4, con 288 genes que preferentemente se encuentran activados en ECs (**Figura R44C**). Finalmente, para corroborar si las EPCs generadas a partir de diferenciación de ESCs obtenidas en este estudio se correspondían transcripcionalmente con las EPCs previamente descritas, procedí en primer lugar a obtener su transcriptoma completo por RNA-seq (**Figura**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

R44A). De modo importante, mediante análisis de componentes principales (PCA) se confirmó que las EPCs generadas en este estudio se agrupaban perfectamente con las EPCs previamente descritas (**Figura R44D**).

En conjunto, mis datos definen por primera vez una firma de expresión génica de especificación endotelial, que permite distinguir entre ESCs, EPCs y ECs. Además, confirman que el protocolo que he optimizado para generar EPCs a partir de ESCs murinas da como resultado células prácticamente indistinguibles transcripcionalmente de sus homólogas.

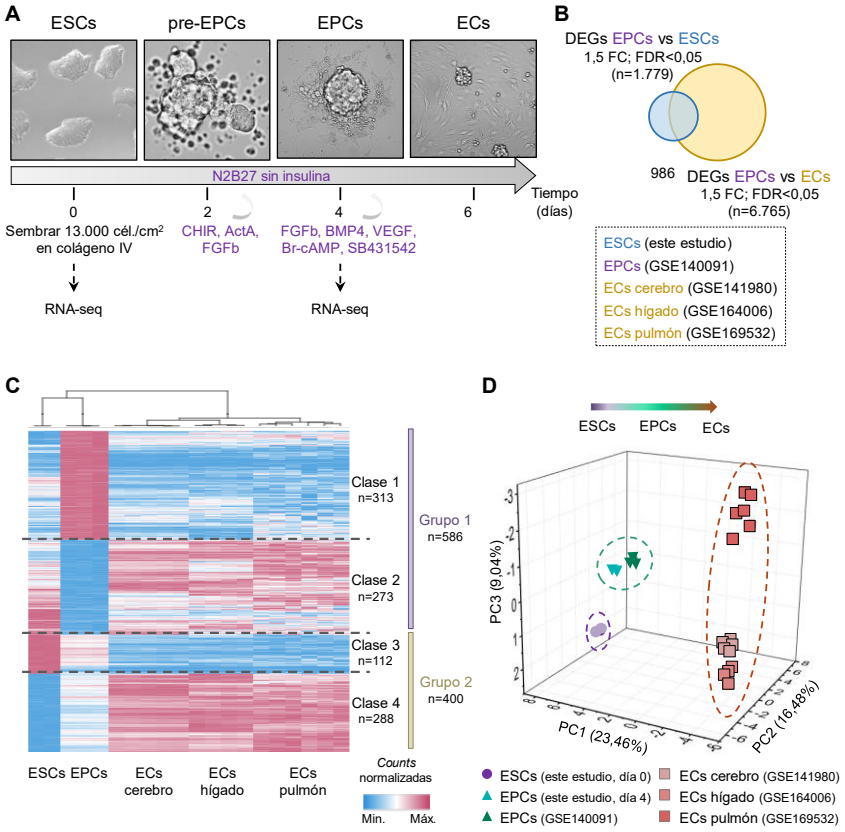


Figura R44. Establecimiento de una firma de expresión génica para EPCs y validación de las EPCs generadas a partir de ESCs. A) Representación gráfica de la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

estrategia seguida para obtener células endoteliales maduras. Se sembraron 13.000 células/cm^2 de ESCs WT en placas tratadas con 18 µg/cm^2 de colágeno IV. Se cultivaron 6 días en medio N2B27 sin insulina. Entre los días 2 y 4 se suplementó el medio con CHIR, ACTIVINA A y FGFb. Entre los días 4 y 6 se suplementó el medio con FGFb, BMP4, VEGF, Br-cAMP y un inhibidor de los receptores tipo I ALK4/5/7 de la superfamilia TGFβ (SB431542). Se muestran imágenes representativas de ESCs, pre-EPCs, EPCs y ECs por microscopía visible. Se indican los tiempos en los que se extrajo ARN para analizar mediante secuenciación masiva (RNA-seq). **B)** Representación mediante diagrama de Venn de los genes diferencialmente expresados (DEGs) entre EPCs vs ESCs y EPCs vs ECs mostrando los genes comunes entre las dos comparaciones. Se emplearon datos de RNA-seq generados en esta tesis y en los estudios de Furuta y colaboradores (GSE164006)[303], Harland y colaboradores (GSE140091)[304], Marchetti y colaboradores (GSE141980)[305] y Rahman y colaboradores (GSE169532)[306]. Los 986 genes comunes constituyen la firma génica de las EPCs. **C)** Representación mediante mapa de calor de la firma génica de especificación endotelial acoplado a una agrupación jerárquica. Se emplearon los datos de *counts* normalizadas para todos los replicados de RNA-seq. Se muestran genes cuya expresión máxima o mínima se produce de manera significativa y con un cambio superior al 50% en las EPCs (grupo 1) o en ESCs/ECs (grupo 2). Dentro de cada grupo se ordenaron los genes en clases dependiendo de su represión o activación. **D)** Análisis PCA de los perfiles transcripcionales de las ESCs y EPCs generadas en esta tesis y las EPCs y ECs de los estudios descritos en **B**. Se utilizaron los 986 genes que constituyen la firma de expresión génica elaborada. El porcentaje de variación explicado por cada componente principal (PC) se indica en el eje correspondiente.

5.3.5 Las proteínas CCM son necesarias para generar EPCs

Tras comprobar y validar la identidad de las células progenitoras endoteliales (EPCs) obtenidas en este trabajo (**Figura R44**), me propuse evaluar la función CCM durante el proceso de especificación endotelial a partir de ESCs para dar lugar a EPCs en cultivo. Para ello, se sembraron ESCs WT y KO para *Ccm1*, *Ccm2* o *Ccm3* en colágeno IV y se mantuvieron 2 días en medio N2B27 sin insulina, seguidos de 2 días en medio suplementado con CHIR99021, ACTIVINA A y FGFb. Se recogieron muestras de ARN a día 0 y día 4 para su análisis mediante secuenciación masiva (RNA-seq) (**Figura R45A**). El análisis mediante PCA, utilizando los valores transcripcionales de los 986 genes que constituyen la firma de expresión génica para EPCs, reveló que la ausencia de CCM1, CCM2 o CCM3 era suficiente para alterar los patrones de expresión dinámicos que conducen a las EPCs WT (**Figura R45B**). En particular, se vio que: i) de manera similar a lo que sucede

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

en las ESCs KO para *Ccm2*, la pérdida de CCM2 daba lugar a células diferenciadas con un perfil transcripcional bastante alejado de las EPCs generadas a partir de ESCs WT y ii) la ausencia de CCM3, con respecto a la pérdida de CCM1, tenía un mayor efecto a nivel de separación transcripcional dificultando la generación de "verdaderas" EPCs (**Figura R45B**). Por otro lado, la comparación de las EPCs WT con las células diferenciadas procedentes de las ESCs KO para cada gen *Ccm*, mostró cambios de expresión significativos tanto al alza como a la baja en ausencia de CCM1, CCM2 y CCM3, siendo el efecto de la pérdida de CCM3 el que más alteró este paisaje transcripcional (**Figura R45C**). Luego, para explorar si estas alteraciones de expresión podrían comprometer el proceso de especificación endotelial, realicé un análisis mediante GSEA de una lista de genes involucrados en la diferenciación celular endotelial (GO:0045446). De manera destacada, se observó que estos genes se encontraban enriquecidos en las células KO para *Ccm1* y *Ccm3* diferenciadas a día 4 en comparación a las EPCs WT, mientras que en el caso de *Ccm2* KO sucedía lo contrario (**Figura R45D**). Dado que la correcta especificación de las células progenitoras endoteliales influye en su capacidad para dar lugar a células endoteliales maduras y en el desarrollo de una función endotelial adecuada, me propuse evaluar mediante análisis GSEA el efecto de la eliminación de las CCM en vías y procesos relevantes para la homeostasis y la función de las células endoteliales. Así, se detectó que las células diferenciadas a día 4 procedentes de las ESCs KO para *Ccm1*, *Ccm2* y *Ccm3* mostraban vías y procesos significativamente afectados de manera común relacionados con la angiogénesis, desarrollo de vasos sanguíneos, adhesión celular y fibras de estrés entre otros (**Figura R45E**). Además, algunas de estas vías se encontraban diferencialmente enriquecidas con respecto a las EPCs WT entre las *Ccm2* KO y las resultantes de la eliminación de CCM1 y CCM3 (por ejemplo, vías de VEGFR, TGFBR y RAP1 o el proceso relacionado con desarrollo de ECs) (**Figura R45E**).

Estos resultados muestran que la presencia individual de cada proteína CCM resulta importante para la generación de "verdaderas"

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

EPCs durante diferenciación, y sugieren la posible existencia de mecanismos moleculares comunes y específicos a cada proteína CCM durante el proceso de especificación endotelial.

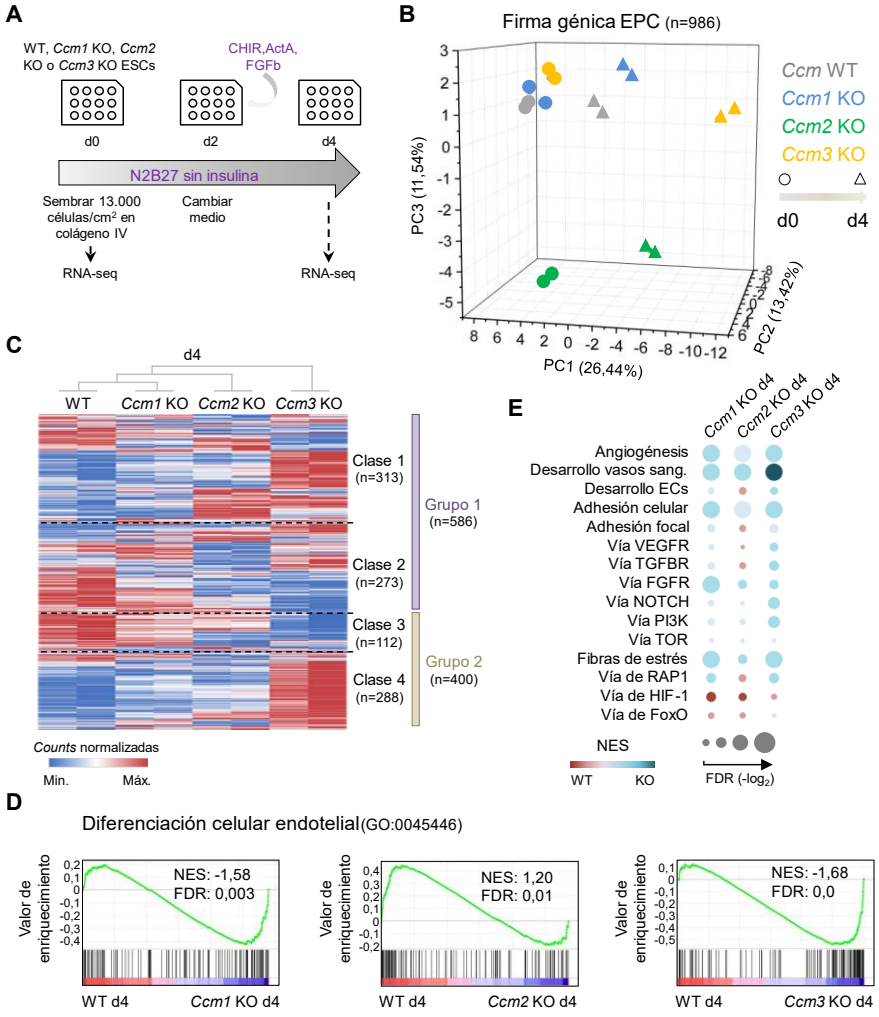


Figura R45. La ausencia de las CCM impide una correcta formación de EPCs. A) Representación gráfica de la estrategia seguida para obtener EPCs. Se sembraron 13.000 células/cm² de ESCs WT o KO para *Ccm1*, *Ccm2* o *Ccm3* en placas tratadas con 18 µg/cm² de colágeno IV. Se cultivaron 4 días en medio N2B27 sin insulina. Entre los días 2 y 4 se suplementó el medio con CHIR, ACTIVINA A y FGFb. Se indican los tiempos en los que se extrajo ARN para analizar mediante secuenciación masiva

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(RNA-seq). **B)** Análisis PCA de los perfiles transcripcionales de las células WT o KO para *Ccm1*, *Ccm2* o *Ccm3* a día 0 y día 4. Se empleó la firma de expresión génica para EPCs generada en este estudio. El porcentaje de variación explicado por cada componente principal (PC) se indica en el eje correspondiente. **C)** Representación mediante mapa de calor de la expresión de los genes que conforman la firma de expresión génica para EPCs acoplado a una agrupación jerárquica. Se emplearon los datos de *counts* normalizadas de RNA-seq para las EPCs WT o células diferenciadas KO para *Ccm1*, *Ccm2* o *Ccm3* a día 4. **D)** Análisis GSEA de los datos de RNA-seq de EPCs WT y células diferenciadas KO para *Ccm1* (izquierda), *Ccm2* (centro) y *Ccm3* (derecha) contra la lista de genes involucrados en la diferenciación celular endotelial (GO:0045446). Se muestra el valor de NES y FDR. **E)** Representación mediante gráfica de burbujas de análisis GSEA de los datos de RNA-seq de EPCs WT y células diferenciadas KO para *Ccm1*, *Ccm2* y *Ccm3* contra las listas indicadas de genes involucrados en procesos y vías de señalización relevantes para la homeostasis y la función endotelial. El color es dependiente del valor del NES mientras que el tamaño lo es del valor de FDR ($-\log_2$).

5.3.6 La pérdida de cada CCM altera genes comunes relacionados con la función endotelial

Para tratar de entender mejor los mecanismos moleculares afectados por la pérdida de las proteínas CCM durante la especificación de células progenitoras endoteliales, analicé el transcriptoma en cada condición KO tras 4 días de diferenciación para caracterizar el papel individual de CCM1, CCM2 y CCM3. Con este fin, clasifiqué como genes diferencialmente expresados (DEGs) todos aquellos que presentaban cambios en su expresión superiores al 50% (>1,5 FC) y con una tasa de descubrimientos falsos inferior al 0,05 (FDR<0,05). Mediante este análisis pude identificar que la ausencia de CCM1 provocaba la desregulación de 800 genes comparado con las EPCs WT, mientras que la falta de CCM2 y CCM3 alteraban la expresión de 1.998 y 4.023 genes, respectivamente (**Figura R46A**). Además, en cada uno de los transcriptomas alterados, la mayoría de los genes afectados (95% en *Ccm1* KO, 87% en *Ccm2* KO y 91% en *Ccm3* KO) codificaban para proteínas (**Figura R46A**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

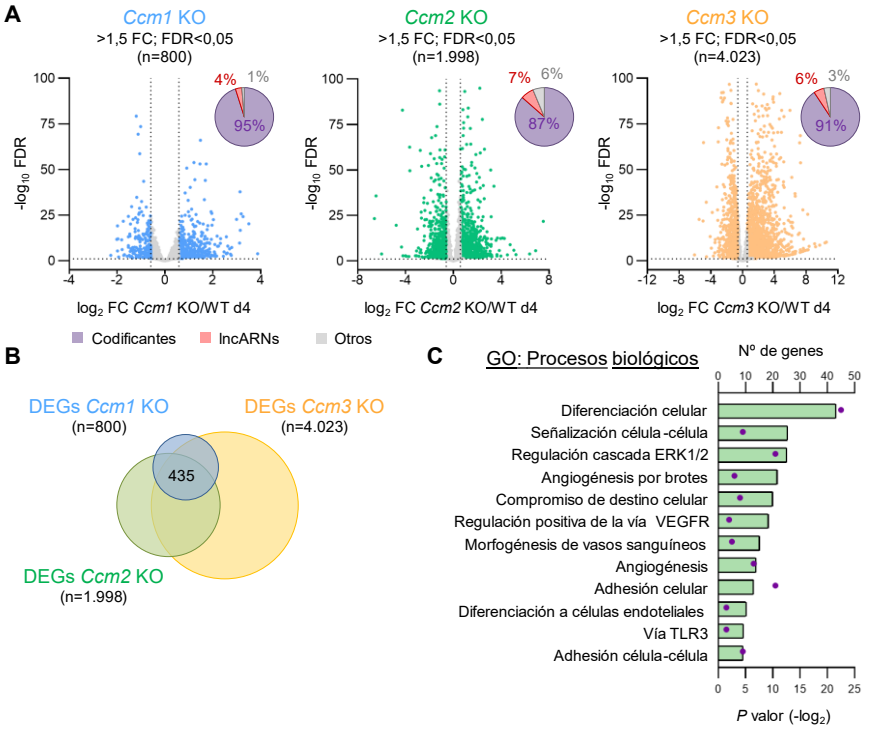


Figura R46. Identificación de genes comunes desregulados en ausencia de cada proteína CCM relacionados con la función endotelial. A) Representación mediante "*volcano plot*" de la expresión génica en *Ccm1* KO (izquierda), *Ccm2* KO (centro) y *Ccm3* KO (derecha) empleando los datos de RNA-seq. Se resaltan los genes diferencialmente expresados (DEGs) entre cada KO y WT (>1,5 FC y FDR<0,05). Se muestra el porcentaje de transcritos desregulados que constituyen ARNs codificantes de proteínas, ARNs largos no codificantes y otros ARNs para cada condición. **B)** Representación mediante diagrama de Venn de los DEGs en las células KO para *Ccm1*, *Ccm2* y *Ccm3* a día 4 tras diferenciación mostrando los genes comunes en las tres condiciones. **C)** Análisis de ontología génica (GO) de procesos biológicos empleando los 435 genes desregulados. Se muestra el enriquecimiento en diversos términos a través de la significancia y el número de genes involucrados en cada proceso. Los términos representados se seleccionaron entre los más enriquecidos.

Dado que la pérdida de función de cualquiera de los tres genes *Ccm* causa defectos en la función endotelial que desencadenan la patología de malformaciones cavernosas cerebrales[220-224], me propuse evaluar las alteraciones transcripcionales que son comunes a la pérdida

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de cada una de las tres proteínas CCM. Para ello, comparé la lista de genes diferencialmente expresados en cada *Ccm* KO tras 4 días de diferenciación, lo que me permitió identificar 435 genes desregulados que presentan la misma tendencia de activación o represión comparada con las EPCs WT (**Figura R46B**). De manera consecuente, al realizar un análisis de ontología génica de estos 435 genes, se encontró que una gran parte de los procesos biológicos alterados de manera significativa se relacionaban con la biología endotelial (por ejemplo, procesos de angiogénesis, morfogénesis de vasos sanguíneos, de unión y señalización entre células) y procesos de diferenciación y especificación de células endoteliales (**Figura R46C**), consistente con los resultados mostrados anteriormente. De manera importante, la eliminación de *Ccm1*, *Ccm2* y *Ccm3* durante la generación de EPCs altera la expresión de genes que intervienen en la regulación de la cascada de ERK1/2 (**Figura R46C**), de manera similar a mis resultados obtenidos en el contexto de mantenimiento de pluripotencia para las ESCs KO de *Ccm2* (**Figura R22**).

Con el fin de analizar cómo la ausencia de cada una de las proteínas CCM puede mediar una desregulación transcripcional común, determiné que factores de transcripción se unen potencialmente a las regiones promotoras de los 435 genes afectados. Para ello, tomé ventaja del software HOMER (del inglés *hypergeometric optimization of motif enrichment*), que comprende una agrupación de herramientas para realizar diferentes análisis bioinformáticos sobre conjuntos de datos de secuenciación masiva. Empleando esta aproximación *in silico*, realicé un análisis de descubrimiento de motivos consenso entre los genes desregulados. De esta manera, se buscaron motivos de unión conocidos de una longitud de 6-12 pares de bases en una ventana de 4 kb alrededor del sitio de inicio de la transcripción (TSS) entre los 435 genes desregulados comunes a la pérdida de función de cada CCM en las células diferenciadas hacia EPCs. Se encontraron 39 motivos de factores de transcripción conocidos, de los cuales 19 están relacionados

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

con distintos aspectos de la biología endotelial, destacando el proceso de angiogénesis (**Tabla R1**).

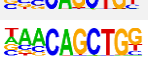
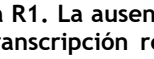
Motivo consenso	Factor	*P* valor	Relacionado con	Referencia
	SP2	0,001	Migración endotelial	Das *y cols.*, 2006[307]
	IRF1	0,001	Angiogénesis inducida por VEGF	Lee *y cols.*, 2008[308]
	SMAD4	0,01	Integridad vascular	Lan *y cols.*, 2007[309] Li *y cols.*, 2011[310]
	TEAD1	0,01	Angiogénesis	Mammoto *y cols.*, 2018[311]
	TCF12	0,01	Angiogénesis	Yang *y cols.*, 2019[312] Fan *y cols.*, 2018[313]
	KLF14	0,01	Función endotelial	Hao *y cols.*, 2018[314]
	TCF21	0,01	Angiogénesis	Duan *y cols.*, 2019[315]
	E2A	0,01	Angiogénesis	Tanaka *y cols.*, 2010[316]
	STAT4	0,01	Vasculogénesis / Angiogénesis	Meng *y cols.*, 2017[317] Tuli *y cols.*, 2022[318]
	P53	0,01	Estado redox	Akhter *y cols.*, 2020[319]
	TEAD4	0,01	VEGF	Appukuttan *y cols.*, 2012[320]
	E2F3	0,01	Proliferación endotelial	Ding *y cols.*, 2020[321] Zhou *y cols.*, 2013[322]
	AP4	0,01	Función endotelial	Burnstock, 2008[323]
	SMAD2	0,01	Angiogénesis / Integridad vascular	Ding *y cols.*, 2020[324] Itoh *y cols.*, 2012[325]
	MAZ	0,01	Angiogénesis inducida por VEGF	Smits *y cols.*, 2012[326]
	P63	0,01	VEGF	Senoo *y cols.*, 2002[327]
	SIX4	0,01	Angiogénesis	Sun *y cols.*, 2019[328]
	LEF1	0,01	Especificación de destino endotelial	Hübner *y cols.*, 2017[329]
	GATA2	0,01	Integridad de unión celular	Mahamud *y cols.*, 2019[330]

Tabla R1. La ausencia de las CCM altera la expresión de genes diana de factores de transcripción relacionados con la función endotelial. Representación de los motivos encontrados en las regiones promotoras de los 435 genes desregulados en las células KO para *Ccm1*, *Ccm2* y *Ccm3* diferenciadas hacia EPCs relacionados con la función endotelial. Se empleó el software HOMER definiendo una ventana de 4 kb

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

alrededor del TSS. Se muestra el factor de transcripción asociado a cada motivo, el *P* valor, la relación descrita con células endoteliales y sus referencias bibliográficas con respecto a la función indicada.

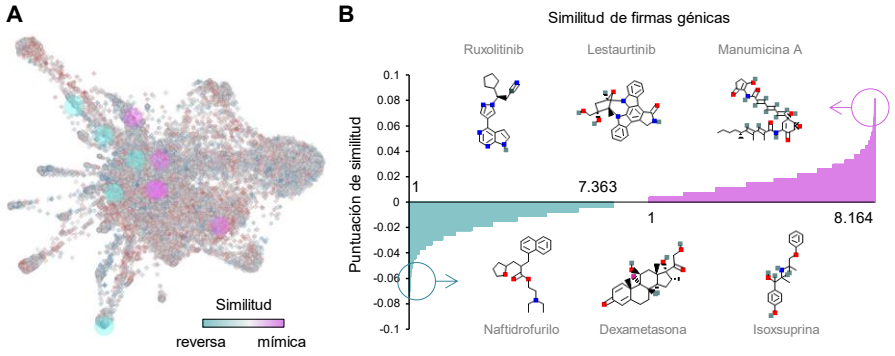
Como se ha visto anteriormente, en ESCs la falta de CCM2 causa un bloqueo en la activación de ERK1/2, provocando alteraciones en los procesos de cambio de identidad celular. En conjunto, mis resultados indican que la ausencia de las proteínas CCM altera el paisaje transcripcional durante la especificación de células progenitoras endoteliales, comprometiendo la generación de estas células y la función endotelial. Además, en línea con los efectos observados tras la eliminación de CCM2 en ESCs, los datos presentados sugieren que la alteración de la cascada de ERK1/2 constituye un factor importante detrás de los procesos de diferenciación y cambio de identidad celular aberrantes.

5.3.7 Identificación de intervenciones farmacológicas potenciales para modular la ausencia de función CCM en el contexto de especificación endotelial

Como se ha mencionado a lo largo de este trabajo, las mutaciones de pérdida de función en cualquiera de los genes *Ccm* altera la función endotelial y causa malformaciones cavernosas cerebrales, una enfermedad para la que actualmente solo existen tratamientos paliativos. Por esta razón, me propuse emplear los datos obtenidos de la especificación de células progenitoras endoteliales a partir de ESCs para identificar potenciales intervenciones farmacológicas que pudiesen neutralizar las alteraciones transcripcionales comunes causadas por la ausencia de cada una de las proteínas CCM. Con este fin, tomé ventaja de la herramienta online L1000FWD (https://maayanlab.cloud/L1000FWD/), donde se recogen las firmas de expresión génica que generan más de 16.000 fármacos y pequeñas moléculas[256]. Para la realización de este análisis, es importante discernir entre los genes desregulados reprimidos y activados. En el caso de los 435 genes diferencialmente expresados tras la eliminación de *Ccm1*,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Ccm2 o *Ccm3* en la diferenciación hacia EPCs, 126 correspondían a genes con expresión reducida y 309 a genes con expresión aumentada (FC>1,5 y FDR<0,05). Mediante la herramienta L1000FWD se comparó esta firma génica CCM con las firmas asociadas a distintos fármacos y pequeñas moléculas, dando como resultado aquellas intervenciones que conllevan un nivel de similitud mimética o reversa en los cambios transcripcionales de los patrones de expresión génica resultantes de la pérdida de función común CCM (**Figura R47A**).



C

Fármaco	Similitud	Descripción
Trichostatina A	-0,074	Inhibidor de HDAC clase I y II
Bisindolilmaleimida	-0,07	Inhibidor de PKC
Dacinostato	-0,066	Inhibidor de HDAC
Immepip	-0,066	Agonista receptor histomana H_3
Prednisolona	-0,063	Corticoesteroide
Fenretidina	-0,063	Derivado sintético retinoico
Pifithrin-mu	-0,059	Inhibidor de p53
Naftidrofurilo	-0,059	Agonista inverso del receptor 5-HT2A
Dexametasona	-0,059	Glucocorticoide sintético
Isoxsuprina	-0,055	Vasodilatador

D

Fármaco	Similitud	Descripción
Alvespimicina	0,081	Inhibidor selective de Hsp90
Rusovastatina	0,07	Inhibidor de la reductase HMG-CoA
AG-14361	0,07	Inhibidor de PARP1
Idarubicina	0,066	Interfiere con la topoisomerase II
Forskolina	0,066	Activa la adenilil ciclasa
Daunorrubicina	0,066	Interfiere con la topoisomerase II
Tetrindol	0,063	Inhibidor de la monoaminooxidasa A
Ruxolitinib	0,063	Inhibidor de JAK
Lestaurtinib	0,063	Inhibidor de JAK2 y TRKA
Manumicina A	0,063	Inhibidor de la farnesiltransferasa

Figura R47. Identificación de fármacos que reviertan los defectos de las EPCs. A) Representación gráfica del efecto de 16.848 fármacos y su similitud con la desregulación génica común a EPCs KO para *Ccm1*, *Ccm2* y *Ccm3* generada con la herramienta L1000FWD. Se destaca la localización de los 5 fármacos con mayor

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

efecto opuesto (azul) y los 5 fármacos más similares (rosa). **B)** Representación mediante gráfico de barras del valor de similitud de la firma génica generada por cada fármaco con los defectos comunes de las EPCs KO. Se representan 7.363 fármacos que revierten estos defectos y 8.164 que los mimetizan. Se muestra la estructura molecular de 3 de los fármacos con mayor efecto de cada grupo. **C)** Lista con los 10 fármacos con los efectos más opuestos, mostrando su puntuación de similitud y su función asociada. **D)** Lista con los 10 fármacos con los efectos más similares, mostrando su puntuación de similitud y su función asociada.

De manera destacada, se halló que entre las intervenciones farmacológicas que presentaban efectos transcripcionales opuestos a los detectados en las células diferenciadas KO para cada gen *Ccm*, se encontraban el naftidrofurilo, la dexametasona y la isoxsuprina, que son empleados en el tratamiento de otras enfermedades y dolencias vasculares (**Figura R47B**). Asimismo, mediante esta aproximación *in silico* se identificaron las principales intervenciones farmacológicas que podrían revertir (**Figuras R47B** y **R47C**) y mimetizar (**Figuras R47B** y **R47D**) el efecto de la pérdida de función de los genes *Ccm* en el contexto de la biología endotelial. Además, estos resultados proporcionan una nueva oportunidad para continuar explorando los mecanismos moleculares en los que intervienen las proteínas CCM en la función endotelial y en la patología asociada a su ausencia.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6. DISCUSIÓN

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Durante las últimas dos décadas, numerosas evidencias científicas han contribuido a un mejor entendimiento del papel de las proteínas CCM en la biología endotelial y en la respuesta de estas células ante su ausencia en el contexto patológico de las malformaciones cavernosas cerebrales. Además, se ha visto como la ausencia de estas proteínas causa defectos morfológicos visibles durante el desarrollo embrionario alrededor de los días E8.5 y E10, afectando el mantenimiento y la formación de redes vasculares, y causando en última instancia la pérdida de viabilidad de los embriones. Sin embargo, no se ha analizado en profundidad el papel de estas proteínas durante el desarrollo embrionario temprano, en el que se establece la pluripotencia celular, ni en los cambios de identidad celular que determinan la especificación hacia células endoteliales. En el transcurso de mi tesis doctoral, busqué definir el papel individual de cada proteína CCM en el contexto de pluripotencia abarcando las capacidades únicas que poseen estas células de autorrenovación ilimitada en cultivo y de dar lugar a cualquier tipo celular de un organismo. Por medio del uso de células troncales embrionarias (ESCs) de origen murino analicé los efectos de la pérdida de CCM1, CCM2 y CCM3. Los datos obtenidos indican que estas proteínas no son imprescindibles para mantener un estado pluripotente celular en cultivo, pero sí para un correcto paisaje transcripcional en condiciones estándar de pluripotencia, afectando incluso la expresión de genes relacionados con funciones endoteliales. En particular, la ausencia de CCM2 promueve la reprogramación de las ESCs de un estado celular pluripotente heterogéneo hacia un estado celular con características similares a la pluripotencia *naïve*, que se corresponde con la pluripotencia *in vivo* de la etapa pre-implantacional uterina del embrión. Mis datos muestran que, a través de su red de proteínas interactoras, CCM2 participa de manera directa en el control epigenético de la expresión de un conjunto de genes relacionados con la regulación de estados alternativos de pluripotencia (p. ej. *naïve* y *primed*) entre los que se encuentran moduladores de la activación de la vía de señalización ERK1/2, cuya activación se encuentra comprometida en las ESCs KO para *Ccm2*. De manera indirecta, la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ausencia de CCM2 conlleva la alteración del control epigenético sobre el paisaje transcripcional pluripotente ejercido por el complejo represor PRC2 y las enzimas reguladoras de la metilación genómica (DNM3B/L, TET1 y TET2). Como consecuencia del control ejercido por CCM2 de manera directa e indirecta en el contexto pluripotente, su ausencia conlleva la retención de un estado pluripotente *naïve* que dificulta el establecimiento y/o mantenimiento del estado pluripotente más avanzado en desarrollo tras la implantación del embrión denominado como *primed* (**Figura D1A**).

Por otro lado, me propuse analizar la relevancia de las proteínas CCM en los procesos de salida del estado celular pluripotente y, en particular, debido a su relevancia en la biología endotelial y enfermedad, durante la especificación celular para dar lugar a células progenitoras endoteliales (EPCs). Los resultados obtenidos revelaron que la ausencia de CCM1, CCM2 y CCM3 compromete la diferenciación celular de ESCs. Además, empleando un protocolo de diferenciación a células endoteliales optimizado durante el transcurso de mi tesis, se pudo comprobar que cada proteína CCM resulta clave para la correcta generación de EPCs, a través del control preciso de la expresión de reguladores de la función endotelial (**Figura D1B**).

En su conjunto, los mecanismos moleculares en los que participan las CCM identificados a lo largo de mi tesis doctoral contribuyen a un mejor entendimiento de los complejos procesos de toma de decisiones celulares que intervienen en desarrollo y cuya alteración puede causar procesos fisiopatológicos como es el caso de las malformaciones cavernosas cerebrales.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

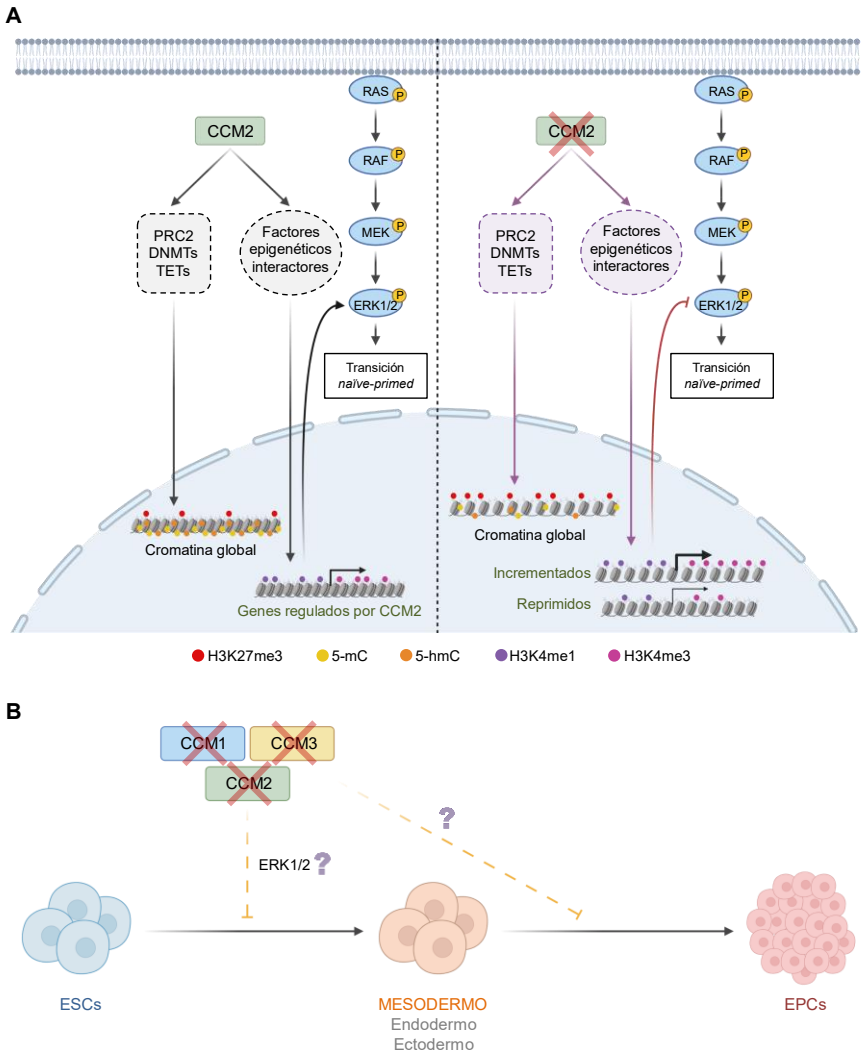


Figura D1. Representación esquemática simplificada de la relevancia de las proteínas CCM en pluripotencia y especificación endotelial. (A) En las ESCs, en condiciones estándar de cultivo (SL), CCM2 a través de su red de interactores interviene en el control de la deposición de marcas epigenéticas activadoras (H3K4me1 y H3K4me3) en regiones potenciadoras y promotoras sobre un conjunto de genes entre los que se incluyen reguladores de la vía de ERK1/2. La ausencia de CCM2 deriva en la falta de activación de ERK1/2, desencadenando un estado pluripotente similar al *naïve*, caracterizado por un aumento en los niveles globales de H3K27me3 e hipometilación a nivel genómico, comprometiendo la transición al estado

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

pluripotente *primed* y afectando a la capacidad de diferenciación de las ESCs. **(B)** Las proteínas CCM son necesarias para la correcta especificación de las tres capas germinales (mesodermo, endodermo y ectodermo), necesaria para que se produzca de manera normal la gastrulación y la formación de células progenitoras endoteliales (EPCs), a través de programas moleculares que no se han identificado, pero en los que podría tener un papel relevante la vía de señalización ERK1/2. Imagen creada con BioRender.com.

6.1 PROTEÍNAS CCM EN ENFERMEDAD

Como se ha descrito a lo largo de las diferentes secciones de esta tesis, las proteínas CCM están involucradas en el desarrollo de las redes vasculares y en el mantenimiento de la arquitectura de los vasos sanguíneos, y su pérdida individual causa malformaciones cavernosas cerebrales mediante la alteración de la función endotelial. Esta enfermedad se da principalmente en la vasculatura del sistema nervioso central, a pesar de expresarse las proteínas CCM de forma ubicua a través de varios tejidos (**Figura R1)**. Se manifiesta así la relevancia de conocer el papel de estas proteínas en contextos diferentes al endotelial para entender las potenciales interconexiones que puedan condicionar la aparición de mutaciones somáticas de pérdida de función en los genes *Ccm* o la reacción adversa a su pérdida. Se desconocen los factores que provocan el paso de lesiones quiescentes a lesiones maduras y activas. El número, tamaño y agresividad de las lesiones CCM varían ampliamente entre pacientes de edades similares con la misma mutación o incluso dentro de miembros de la misma familia. Esto hace pensar que el ambiente y las condiciones fisiológicas particulares de cada individuo puedan condicionar la respuesta celular y el efecto de la ausencia de las proteínas CCM. Por ejemplo, se ha visto que, si bien la mayoría de los casos esporádicos ocurren en personas sanas, la radiación ionizante es un potente inductor de la formación de cavernomas[331-333]. Por otro lado, se ha reportado que otros factores ambientales, como el microbioma y la barrera intestinal, pueden fomentar la formación de carvernomas[334,335]. Teniendo estos estudios en mente, resultaría interesante estudiar qué factores externos son capaces de incrementar la ratio mutacional de los genes *Ccm*. Para ello,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

emplearía los clones heterocigotos generados durante el proceso de obtención de líneas celulares troncales embrionarias KO para *Ccm1*, *Ccm2* o *Ccm3*. Sobre estas células generaría una línea reportera fluorescente (por ejemplo, GFP o mCherry) mediante el sistema CRISPR-Cas9 bajo el control del alelo *Ccm* funcional. Esto me serviría, mediante el empleo de cribados masivos con estímulos físicos y/o químicos, para establecer una nueva estrategia en el campo de estudio de esta patología y poder interrogar qué clase de estrés celular fomenta la aparición de mutaciones somáticas de pérdida de función en los genes *Ccm*.

Entre la gran cantidad de funciones celulares que se han atribuido a las proteínas CCM, algunas siguen siendo controvertidas, por lo que es necesario esclarecer las interacciones y/o las vías de señalización deterioradas que desencadenan la formación de cavernomas. Al contrario de lo observado en células endoteliales, mis resultados muestran que las proteínas CCM no son necesarias para mantener la autorrenovación de las células pluripotentes en presencia de suero (**Figuras R3-R5**). A pesar de ello, incluso en este contexto, su presencia es necesaria para la correcta regulación de la expresión de genes involucrados en procesos y vías relacionadas tanto con la pluripotencia como con la función endotelial. Mis resultados muestran que en las ESCs no solo se ven afectados procesos como la angiogénesis y la adhesión celular, sino procesos de diferenciación (**Figuras R8, R18, R19, R28, R36** y **R40**), lo que apunta a una capacidad comprometida para determinar tanto un destino endotelial como aquellos destinos relacionados con los distintos nichos vasculares. Consecuentemente, los datos obtenidos muestran procesos de diferenciación aberrantes tanto en ensayos de diferenciación a cuerpos embrioides en la determinación de las tres capas germinales como en ensayos de formación de gastruloides ante la ausencia de cualquiera de las proteínas CCM (**Figuras R41** y **R42**). Además, la eliminación de estas proteínas altera la especificación de las células progenitoras endoteliales (**Figuras R45** y **R46**), linaje derivado del mesodermo[159].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Esta especificación aberrante provoca la alteración de procesos y vías de señalización esenciales para la función endotelial (**Figuras R45, R46** y **Tabla R1**). Una de las principales vías afectadas es la señalización por VEGF, crítica para una correcta especificación endotelial y para la regulación del proceso de angiogénesis[336,337]. Curiosamente, se ha mostrado que la inhibición de esta vía de señalización resulta ser beneficiosa para mantener el estado de autorrenovación de las células troncales embrionarias durante su cultivo a largo plazo[338,339]. Por otro lado, la ausencia de las proteínas CCM también altera la vía de HIF-1, que además de ser importante para la regulación de la homeostasis de las especies reactivas de oxígeno (ROS), se ha mostrado que desempeña un papel crucial en el mantenimiento de las ESCs o en la aceleración de su diferenciación hacia el linaje vascular[340,341].

Comprender en profundidad los mecanismos subyacentes de la enfermedad puede ayudar a determinar aproximaciones farmacológicas que puedan prevenir que las lesiones se vuelvan sintomáticas o incluso que detengan su progresión. La especificación de células progenitoras endoteliales llevada a cabo durante esta tesis me permitió analizar las alteraciones transcripcionales comunes en ausencia de CCM1, CCM2 o CCM3 y definir una lista de posibles intervenciones farmacológicas para revertir, o al menos disminuir, estas variaciones (**Figura R47**). Como se ha mencionado anteriormente, la pérdida de función de las proteínas CCM en células progenitoras endoteliales es capaz de iniciar lesiones cavernosas[236,237], por lo que la modificación de las alteraciones generadas en estas células puede resultar de gran relevancia en el contexto de la enfermedad CCM. Resultaría interesante realizar un cribado con los fármacos y pequeñas moléculas identificados en este trabajo, para analizar en profundidad el grado de reversión que provocan en la firma de expresión génica EPC alterada en ausencia de cada gen *Ccm* y sus posibles efectos secundarios en la biología endotelial. Además, para paliar mejor las alteraciones dependientes de cada gen *Ccm* individual, convendría identificar fármacos que puedan

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

oponerse a cada firma génica concreta y no solo a las variaciones comunes. Finalmente, sería necesario analizar la capacidad de las células progenitoras endoteliales tratadas con los compuestos candidatos para dar lugar a células endoteliales maduras funcionales que puedan formar redes tubulares y redes de vasos sin defectos.

6.2 PROTEÍNAS CCM EN DESARROLLO

Durante el desarrollo embrionario murino, tras la formación del mesodermo temprano, una red vascular primitiva comienza a formarse en el saco vitelino extraembrionario proximal alrededor del día embrionario E6.5[336]. Las células mesodérmicas dan lugar a precursores endoteliales que formarán islas sanguíneas con células endoteliales en la periferia y células hematopoyéticas en el centro[336]. La fusión de estas islas da lugar a una red vascular simple, todo ello en ausencia de circulación sanguínea[336]. Durante la formación de la red vascular extraembrionaria, también tiene lugar la formación de su contraparte intraembrionaria, donde los precursores endoteliales forman la aorta dorsal y la vena cardinal, así como las arterias y venas vitelinas[336]. Alrededor del día E8.0 se establece un sistema circulatorio rudimentario que conecta vasos extraembrionarios e intraembrionarios. El establecimiento mediante vasculogénesis de una red vascular primitiva parece ser independiente de la función asociada a las proteínas CCM, puesto que su ausencia no afecta gravemente este proceso según lo reportado. A partir de este momento, comienza a establecerse la circulación sanguínea, induciendo los procesos angiogénicos y la remodelación vascular[342]. Como se mencionó anteriormente, la ausencia de cualquiera de las tres proteínas CCM resulta letal durante el desarrollo embrionario debido a fallos en estos procesos angiogénicos[178,188,217-219].

Mis resultados sugieren que, a pesar de que las diferencias morfológicas y defectos en la circulación y angiogénesis se observan a partir del día E8.0, podrían existir defectos previos y en contextos celulares diferentes. Las células pluripotentes presentan capacidades

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

aberrantes de gastrulación y determinación de las tres capas germinales en cultivo, incluyendo la capa germinal de mesodermo, tras la eliminación de cualquiera de las proteínas CCM (**Figuras R41** y **R42**). Además, estas alteraciones comprometen la especificación de células progenitoras endoteliales (**Figuras R45** y **R46**). En el caso de producirse *in vivo*, la especificación aberrante del mesodermo podría explicar la pérdida de identidad endotelial y los defectos observados en las redes vasculares tanto del saco vitelino como del embrión.

Por otro lado, mis datos apuntan a que la ausencia de las proteínas CCM en las células progenitoras endoteliales altera la regulación de la vía de señalización de ERK1/2 (**Figura R46**). No solo eso, sino que la pérdida de CCM2 en células troncales embrionarias resulta en la inhibición de la activación de ERK1/2 (**Figura R20**) a través de la alteración transcripcional de varios reguladores positivos y negativos (**Figura R22**). La activación de ERK1/2 resulta crítica para los procesos de salida del estado pluripotente y de diferenciación celular[105,151,152]. Además, varios estudios han puesto de manifiesto la relevancia de la vía de ERK1/2 en la función endotelial y los procesos angiogénicos[343,344]. También se ha relacionado la activación de ERK1/2 con la permeabilidad y apoptosis de las células endoteliales microvasculares del cerebro humano o con la respuesta al estrés oxidativo de las células endoteliales microvasculares cardíacas[345,346]. Resultaría interesante analizar los niveles globales de ERK1/2 activado en las células progenitoras endoteliales, y saber si las ESCs en ausencia de CCM2 son capaces de aumentar su activación ante señales específicas. Para ello, cultivaría ESCs en bajas concentraciones de suero durante un corto periodo de tiempo para bloquear la vía RAS-ERK1/2, volviendo al medio de cultivo pluripotente estándar para estimular y analizar su activación.

Como se ha mencionado, la activación de ERK1/2 parece ser un factor importante detrás de los déficits observados tras la eliminación de las proteínas CCM en diferentes procesos de diferenciación y de especificación de células progenitoras endoteliales (EPCs).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Curiosamente, de manera similar a la ausencia de las proteínas CCM, el bloqueo de la señalización de ERK1/2 provoca la muerte embrionaria entre los días E9.5 y E10.5 debido a un desarrollo cardíaco y placentario aberrante, mostrando los ratones KO para *Erk1/2* defectos durante el proceso de especificación del mesodermo[347-351]. Por otro lado, la señalización de FGF promueve la especificación del mesodermo en detrimento del compromiso neural a través de la cascada de ERK1/2 durante la gastrulación[352].

6.3 MECANISMO MOLECULAR DE LA SEÑALIZACIÓN CCM

Mediante sus diversos dominios, las proteínas CCM pueden unirse a diferentes proteínas, a través de las cuales regulan numerosos procesos celulares. Sin embargo, el panorama de las interacciones y las vías de señalización controladas por las proteínas CCM es muy complejo y quedan muchas preguntas por esclarecer. Además, resulta crítico discernir la función de estas proteínas no solo en células endoteliales sino también en distintos contextos con el fin de comprender mejor como se desencadenan las lesiones cavernosas a partir de una visión global del sistema. La identificación de mutaciones en los genes *Ccm* como causa de las incurables malformaciones cavernosas cerebrales derivó en multitud de esfuerzos para caracterizar su función en células endoteliales y las consecuencias de su pérdida. Sin embargo, aún no se conoce si las funciones e interacciones descritas en un contexto endotelial se repiten en otros contextos o identidades celulares en las que se encuentran presentes las proteínas CCM.

Tanto CCM1 como CCM2 y CCM3 se expresan en células troncales embrionarias (ESCs), mostrando una tendencia al alza durante las primeras etapas del desarrollo embrionario (**Figura R1**). Incluso en un estado pluripotente, puede verse como la ausencia de estas proteínas altera procesos y vías de señalización relevantes para una correcta función endotelial (**Figuras R8, R18, R19, R26, R28, R31, R33** y **R36**). Como se ha descrito anteriormente, el espectro pluripotente representa un conjunto de estados celulares distintos que comprende

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

células equivalentes a la masa celular interna del blastocisto (pluripotencia *naïve*) y células equivalentes a diversos momentos del epiblasto post-implantacional (desde pluripotencia formativa hasta pluripotencia *primed*), más preparadas para responder rápidamente a estímulos de diferenciación. *Ex vivo*, los diferentes estados pluripotentes requieren condiciones de cultivo particulares. Mantenidas en condiciones estándar (presencia de suero y LIF), las ESCs constituyen una población heterogénea que contiene células más resistentes a diferenciarse (*naïve*) y células más proclives a ello (*primed*)[33,34]. Para mantener una población homogénea de células *naïve*, se necesita la ausencia de suero y la presencia en el medio de inhibidores de MEK y GSK3β[17,29]. Mis resultados muestran que en ausencia de CCM2, las ESCs en condiciones estándar se desplazan en el espectro pluripotente, generándose una población similar a la *naïve* (**Figuras R9** y **R10**). Además, mis datos sugieren que la eliminación de CCM2 favorece la reprogramación hacia un estado pluripotente *naïve* y dificulta la obtención y mantenimiento del estado pluripotente *primed* (**Figuras R13** y **R15**). A lo largo del presente trabajo identifiqué la alteración de genes reguladores de ERK1/2 (**Figura R22**) como la causa más probable detrás del bloqueo en la activación de ERK1/2 en ausencia de CCM2 (**Figura R20**), puesto que esta proteína no interactúa directamente con ningún componente de la vía RAS-RAF-MEK-ERK1/2 (**Figura R21**). Como se ha mencionado, la activación y actividad de ERK1/2 es crítica para la salida del estado pluripotente y el comienzo de la diferenciación[105,151,152]. Además, recientemente se ha visto que la presencia de ERK1/2 activado en el núcleo es importante para la transición *naïve-primed*[353].

La ausencia de CCM2 en ESCs desencadena alteraciones en la abundancia de diversos componentes del complejo represor de *Polycomb* 2 (PRC2) y su actividad (**Figura R27**), explicando casi el 40% de los cambios transcripcionales observados en ausencia de CCM2 (**Figura R28**). En esta línea, se ha visto que las células *naïve* cultivadas en presencia de los dos inhibidores (MEK y GSK3β) muestran niveles

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

incrementados de los componentes principales de PRC2 y un aumento global en la deposición de H3K27me3[103]. Sin embargo, se ha descrito una reducción de la presencia de PRC2 y su marca asociada en los genes de desarrollo (bivalentes)[34,104,105]. Varios estudios han mostrado una correlación entre ERK1/2 y miembros de PRC2, determinando la activación de ERK1/2 la unión de varios componentes de este complejo represor epigenético a la cromatina[103,105,146]. Los promotores bivalentes (marcados por H3K4me3 y H3K27me3 simultáneamente) son característicos del estado pluripotente *primed*, que requiere la activación de la vía FGF-ERK1/2 para su mantenimiento en cultivo[38,39,98]. Resultaría interesante, tras haber comprobado la incapacidad de las ESCs KO para *Ccm2* de activar ERK1/2 ante estímulos específicos, analizar si esta quinasa presenta una correcta localización subcelular durante la reprogramación a un estado *primed*. Además, analizaría la deposición de H3K4me3 y H3K27me3 en promotores bivalentes comparándola con células WT reprogramadas y células en estado *primed*. Por otro lado, mis resultados sugieren que la regulación alterada de ERK1/2 puede estar detrás de los defectos observados en los procesos de diferenciación, y especialmente en la especificación de células progenitoras endoteliales ante la ausencia de las proteínas CCM (**Figura R46**). La ausencia de activación de ERK1/2 provoca el aumento de varios componentes de PRC2 en ESCs KO para *Ccm2* (**Figura R27**). Curiosamente, se ha observado que altos niveles de expresión de los componentes centrales del complejo PRC2 son capaces de bloquear la diferenciación endotelial y la formación de tubos[354].

Además de los efectos sobre *Polycomb*, la falta de CCM2 provoca alteraciones en la abundancia y actividad de las enzimas DNMTs y TETs, generando un estado de hipometilación a nivel genómico (**Figuras R29** y **R30**). Concretamente, las ESCs KO para *Ccm2* muestran niveles disminuidos de *Dnmt3b*, DNMT3L y TET1, mientras que presentan niveles incrementados de TET2, explicando en conjunto el 40% de los cambios transcripcionales observados en

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ausencia de CCM2 (**Figura R33**). Como se ha mencionado anteriormente, el estado pluripotente *naïve* se caracteriza por una hipometilación global del ADN y una reducción de los niveles de DNMT3A y DNMT3B acompañada de un aumento de TET2[69,70,73], en consonancia con mis resultados. Por otro lado, se ha visto que TET1, la principal enzima TET expresada en los estados pluripotentes formativo y *primed*[73], es responsable de mantener valles de metilación en los promotores de genes de desarrollo para establecer un estado de "super" bivalencia[98]. Además, recientemente se ha descrito la relación existente entre TET1 y PRC2 en el control de la bivalencia en el estado pluripotente[355]. A este respecto, resultaría interesante analizar si el silenciamiento de TET2 o PRC2 y/o el rescate de TET1 son capaces de revertir los efectos causados por ausencia de CCM2 en la obtención y mantenimiento de un estado pluripotente *primed*.

El hecho de que el cultivo de ESCs en medio 2iL, basado en la presencia de inhibidores de MEK y GSK3β, propicie la aparición y sea crítico para el mantenimiento del estado pluripotente *naïve*, caracterizado por altos niveles globales de PRC2 y TET2, y niveles reducidos de DNMT3A y DNMT3B, sugiere que la inhibición de ERK1/2 provocada por la ausencia de CCM2 es el desencadenante del fenotipo observado a lo largo de mis resultados. Sin embargo, serán necesarios futuros estudios para comprender cómo CCM2 es capaz de afectar la activación de ERK1/2 y cómo esto altera los niveles y actividad de los complejos y enzimas mencionadas.

6.4 MODULADORES DE LA SEÑAL CCM

La identificación de proteínas interactoras del complejo de señalización CCM ha permitido caracterizar moduladores de esta vía en un contexto endotelial. Sin embargo, quedan muchas preguntas pendientes con respecto a los eventos que desencadenan la enfermedad, y por qué se da principalmente en el sistema nervioso central. Esto manifiesta la importancia de conocer la identidad de las proteínas capaces de unirse

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

a las CCM para alcanzar una visión más global y un mejor entendimiento de los mecanismos de acción de estas proteínas.

Mis resultados muestran que, en células pluripotentes en condiciones estándar de cultivo, CCM2 se une a 102 proteínas entre las que se incluye CCM1, pero no CCM3 (**Figura R34**). La pérdida de CCM2 no afecta de manera notable a los niveles de expresión de la mayoría de los genes que codifican a sus interactores (**Figura R35**). Sin embargo, mediante la interacción con reguladores epigenéticos, CCM2 podría influir en la transcripción de diversos genes (entre ellos reguladores de la activación de ERK1/2) a través de la alteración en la deposición de marcas activadoras y/o represoras en regiones potenciadoras y promotoras, derivando todo ello en la aparición de un fenotipo pluripotente *naïve* ante su ausencia (**Figuras R38** y **R39**). Además, se ha visto que varios de los interactores identificados en este estudio se relacionan con la actividad de ERK1/2, aunque sea en contextos celulares diferentes, como por ejemplo CAMSAP3[356], USP36[357], STK38[358] o TMOD3[359]. Por otro lado, de manera independiente a la regulación de ERK1/2, varios interactores de CCM2 tienen un efecto potencial directo tanto sobre la adquisición y mantenimiento de un estado pluripotente *naïve* como en la salida de éste (**Figura R37**).

Con respecto a otras proteínas interactoras identificadas durante esta tesis y su relación con los procesos aberrantes observados, se ha visto que PPP1R12A (MYPT1) es la mayor subunidad regulatoria del complejo fosfatasa MLCP, del que depende la reorganización del esqueleto de actomiosina y su grado de contractilidad[360-362]. Además de ser un proceso crítico para la función endotelial y para una correcta angiogénesis, las interacciones entre la actina y la miosina pueden modular los procesos de diferenciación de las células pluripotentes, y el bloqueo de esta interacción provoca un descenso en la especificación de la capa del mesodermo[363]. Como se ha mencionado anteriormente, las células endoteliales derivan de esta capa germinal, por lo que

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

defectos en su especificación alterarán la obtención y características de las células progenitoras endoteliales.

En conjunto, los resultados obtenidos en mi tesis doctoral muestran que las proteínas CCM, en un contexto pluripotente, son capaces de participar en la regulación de la expresión y actividad de factores involucrados en procesos de diferenciación y especificación endotelial. Además, los mecanismos y funciones en las que intervienen las proteínas CCM en los contextos investigados en este trabajo tienen el potencial de contribuir a un mejor entendimiento de las alteraciones moleculares que se encuentran detrás de la patología de las malformaciones cavernosas cerebrales.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7. CONCLUSIONES

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. La función CCM es dispensable para el mantenimiento de pluripotencia, pero necesaria para un correcto control transcripcional.

2. La ausencia de CCM2 facilita pluripotencia *naïve* y restringe pluripotencia *primed*.

3. CCM2 es necesario para la activación de ERK1/2 en ESCs cultivadas en condiciones estándar.

4. ERK1/2 y los reguladores epigenéticos PRC2, DNMT3B y TET1/2 participan en la regulación transcripcional en la que interviene CCM2 en ESCs.

5. CCM2 a través de su red de interactores en ESCs vincula múltiples vías epigenéticas en el control de pluripotencia.

6. La ausencia de función CCM compromete la capacidad de diferenciación celular de las ESCs.

7. Las proteínas CCM son necesarias en la especificación endotelial, comprometiendo la generación de EPCs en su ausencia.

8. Cada CCM participa en el control adecuado de la firma génica de expresión que caracteriza a las EPCs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8. APÉNDICES

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.1 TABLA SUPLEMENTARIA 1. REACTIVOS Y QUÍMICOS

Nombre	Marca	Referencia
4',6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich	D9542
Acetona	Fisher Chemical	A/0560/17
Acrilamida/bisacrilamida	Sigma-Aldrich	A3574
Activina A	Invitrogen	PHC9564
Acutassa	Sigma-Aldrich	A6964
Adenosina	Sigma-Aldrich	A4036
Agua libre de ribonucleasas	Invitrogen	10977-035
Albúmina de suero bovino (BSA)	Fisher Bioreagents	BP1605
Aminoácidos no esenciales	Gibco	11140-050
Ampicilina	Fisher Bioreagents	BP1760
ATP pH7	TCI Chemicals	A0157
BMP4	Gibco	PHC9534
Bolas magnéticas (Dynabeads Protein G)	Invitrogen	1004D
Br-cAMP	Alfa Aesar	J64469
Bromofenol azul	Sigma-Aldrich	B8026
$CaCl_2 \cdot 2H2O$	Fisher Chemical	VL2151000
Cebadores hexaméricos aleatorios	Thermo Scientific	SO142
CHIR99021	Sigma-Aldrich	SML1046
Citidina	Sigma-Aldrich	C4654
Cloroformo	Fisher Chemical	C/4920/15
Cocktail inhibidor de proteasas (PIC)	Sigma-Aldrich	P8340
Colágeno IV	Sigma-Aldrich	C5533
Desoxicolato sódico	Sigma-Aldrich	D6750
Dimetilsulfóxido (DMSO)	Sigma-Aldrich	75-21-8
DTT	Sigma-Aldrich	43816
EDTA	Sigma-Aldrich	E7889
EGTA	Alfa Aesar	J61721
Enzima de restricción *AgeI*	Thermo Scientific	ER1461
Enzima de restricción *BbsI*	Thermo scientific	ER1012
Enzima de restricción *EcoRI*	Thermo Scientific	ER0271
Enzima de restricción *NheI*	Thermo Scientific	ER0971
Enzima de restricción *PacI*	Thermo Scientific	ER2202
Etanol	Sigma-Aldrich	1.08543
Factor Inhibidor de la Leucemia (LIF)	producido en el laboratorio	n/a
Fenol:cloroformo:isoamil alcohol (PCIA 25:24:1)	Affymetrix	75831
FGFb	Gibco	PHG0026
Fibronectina	Millipore	FC010
Formaldehido 37%	Sigma-Aldrich	F8775
Fosfatasa alcalina termosensible FasAP	Thermo Scientific	EF0654
Gelatina	Sigma-Aldrich	G9391
Glicerol	Sigma-Aldrich	G5516

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Glicina	Sigma-Aldrich	G8898
Glucógeno	Thermo Scientific	R0561
Guanosina	Sigma-Aldrich	G6264
HEPES pH 7,9	Cytiva	SH30237.01
HEPES pH 7,9	Fisher Bioreagents	BP299
HEPES-KOH pH 7,3	Fisher Bioreagents	BP299
Higromicina	Corning	30-240-CR
Igepal	Sigma-Aldrich	I3021
Inhibidor de los receptores tipo I de la superfamilia TGFβ ALK4/5/7	Santa Cruz Biotechnology	sc-204265
Isopropanol	Fisher Bioreagents	BP2618
KCl	Fisher Chemical	P/4240/60
Leche en polvo	Lab Scientific	M0841
L-Glutamina	Gibco	25030-081
LiCl	Fisher Chemical	L/2200/65
Ligasa de ADN T4	Thermo Scientific	EL0016
Medio Eagle modificado de Dulbecco alto en glucosa (DMEM)	Corning	10-013-CV
Medio Eagle modificado de Dulbecco con mezcla de nutrientes F-12 Ham (DMEM/F-12)	Sigma-Aldrich	D6421
Medio Neurobasal	Gibco	21103049
β-mercaptoetanol	Sigma-Aldrich	M-6250
Mezcla de dNTP	Thermo Scientific	R0191
Mezcla maestra de polimerasa de ADN DreamTaq Green PCR	Thermo Scientific	K1081
Mezcla maestra de polimerasa de alta fidelidad Platinum SuperFi II PCR	Invitrogen	12368010
$MgCl_2$	Acros Organics	223210010
$MnCl_2 \cdot 4H2O$	Sigma-Aldrich	M8266
NaCl	Fisher Bioreagents	BP358-1
NaFl	Acros Organics	201290250
N-lauroilsarcosina	Sigma-Aldrich	L91150
Ortovanadato	MP Biomedicals	215966410
Paraformaldehido	Sigma-Aldrich	P6148
PBS con Ca^{+2} y Mg^{+2}	Sigma-Aldrich	D8662
PD0325901	Sigma-Aldrich	PZ0162
Penicilina/estreptomicina	Gibco	15070-063
PIPES pH 6.7	Sigma-Aldrich	P6757
pLKO.1 hygro	Addgene	24150
pMDG2	Addgene	12259
PMSF	Santa Cruz Biotechnology	sc-482875
Polibreno	Sigma-Aldrich	H9268
Polietilenimina (PEI)	Sigma-Aldrich	408727

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PowerUp SYBR Green	Applied Biosystems	A25741
pPB-CAG-3xFLAG-pA-pgk-hph	Addgene	48753
Proteinasa K	Thermo Scientific	EO0491
psPAX2	Addgene	12260
Puromicina	Fisher Bioreagents	BP2956
qScript cDNA SuperMix	Quantabio	P/N 84034
Reactivo de Ponceau S	Sigma-Aldrich	P3504
Reactivo TRIzol	Invitrogen	15596018
RiboLock	Thermo Scientific	E00381
Ribonucleasa A	Thermo Scientific	EN0531
SDS	Fisher Bioreagents	BP1311
SpCas9-2A-Puro V2.0	Addgene	62988
Suero fetal bovino (FBS)	Gibco	10270106
SuperScript IV RT	Invitrogen	18090010
Suplemento B-27	Gibco	17504044
Suplemento B-27 sin insulina	Gibco	A1895601
Suplemento N-2	Gibco	17502048
Sustrato para inmunotransferencia ECL Pierce	Thermo Scientific	32106
T4 PNK	Thermo Scientific	EK0032
Tampón de reacción A de la polinucleótido quinasa T4 (PNK)	Thermo Scientific	EK0032
Tampón de reacción NEB	New England BioLabs	B7002S
Tampón fosfato salino (PBS)	Corning	21-031-CV
Tampón salino Tris	Fisher Bioreagent	BP2471-1
Timidina	Sigma-Aldrich	T1895
Tripsina-EDTA	Sigma-Aldrich	T3924
Tris-HCl pH 6,8	Bio Basic	SD8122
Tris-HCl pH 7,5	Fisher Bioreagents	BP1757
Tris-HCl pH 8,0	Alfa Aesar	J60080
Triton X-100	Sigma-Aldrich	T8787
TWEEN 20	Sigma-Aldrich	P1379
Uridina	Sigma-Aldrich	U3003
VEGF	Sigma-Aldrich	V5765
Yoduro de propidio	Molecular Probes	P3566

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.2 TABLA SUPLEMENTARIA 2. KITS

Nombre	Marca	Referencia
5-mC DNA ELISA	Zymo Research	D5325
Centra Puregene Cell Kit	Qiagen	158767
CloneJet PCR Cloning	Thermo Scientific	K1231
E.Z.N.A Total RNA kit 1	Omega Bio-Tek	R6834-02
Ensayo de proteínas BCA Pierce	Thermo Scientific	23227
Fraccionamiento subcelular de proteínas	Thermo Scientific	78840
GenElute Gel Extraction	Sigma-Aldrich	NA1111
GenElute Plasmid Miniprep	Sigma-Aldrich	PLN350
jetPRIME	Polyplus	114-15
Leucocyte Alkaline Phosphatase	Sigma-Aldrich	86R-1KT
PureLink HiPure Plasmid Filter Midiprep	Invitrogen	K210015
Qubit HS Assay	Invitrogen	Q32855
Qubit RNA IQ Assay	Invitrogen	Q33222
Quest 5-hmC DNA ELISA	Zymo Research	D5425
Sistema de tampón de lisis RIPA	Santa Cruz Biotechnology	sc-24948

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.3 TABLA SUPLEMENTARIA 3. ACCESORIOS Y FUNGIBLE

Nombre	Marca	Referencia
0,22 µm EMD Millipore Steriflip	Millipore	SCGP00525
Agitador de balanceo ProBot Rcoker 25	Labnet	S2025-B
Agitador de incubación	Infors HT	Minitron
Agujas de 18G ½	Terumo	8AN1838S1
Agujas de 21G ½	Terumo	8AN2138R1
Baño térmico serie VWB2	VWR	462-0556
Cámara de Neubauer	Hirschmann	8100103
Casetes de diálisis de Slide-A-Lyzer G2, 10K MWCO	Thermo Scientific	87731
Centrífuga refrigerada	MPW	260R
Centrífuga Sorvall	Thermo Scientific	ST 16R
Citómetro	BD Biosciences	FACSCalibur
Espectofotómetro Nanodrop 2000	Thermo Scientific	ND-2000
Filtros Amicon Ultra de 0,5 mL y 3kDa	Sigma-Aldrich	UFC500308
Filtros Amicon Ultra-15	Millipore	UFC903024
Filtros de 0,22 µm	Fisherbrand	15181489
Fluorómetro Qubit	Invitrogen	Q33226
Imán DynaMag-2	Invitrogen	12321D
Incubador de CO2 New Branswick	Eppendorf	Galaxy 170S
Incubadora digital INCU-Line	VWR	390-0384
Jeringuilla	Terumo	197
Lector de placas Mithras	Berthol Technologies	LB 940
Membranas de fluoruro de polivinilideno	Cytiva	10600023
Mezclador vórtex	Fisherbrand	11746744
Microscopio invertido	Leica	DMI6000 B
Microscopio invertido con una cámara Olympus DP72	Olympus	IX51
Mini Gel Tank	Invitrogen	A25977
Novex WedgeWell 4–20% tris-Glycine Gel	Invitrogen	XP04205BOX
Novex WedgeWell 4–20% tris-Glycine Gel (10 pocillos)	Invitrogen	XP04200BOX
Películas de rayos X Fuji	Fujifilm	100NIF
Películas PlateMax UltraClear	Axygen	UC-500
Placa de 12 pocillos	Corning	3513
Placa de 48 pocillos	Corning	3548
Placas de 96 pocillos con fondo en U	Corning	3788
Placas de cultivo de 150 mm	Corning	430599
Placas de reacción de 384 pocillos	Axygen	PCR-384M2-C
Placas Petri de baja adherencia	Corning	3472
Procesador de películas Fuji	Fujifilm	FPM-100A
Qubit assay tubes	Invitrogen	Q32856
Rotador	Grant-bio	PTR-35

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Sistema de electroforesis horizontal kuroGEL	VWR	700-0034
Sistema de electroforesis vertical Mini-PROTEAN Cell	Bio-Rad	1658004
Sistema de filtrado de 1 L	Corning	431098
Sistema de filtrado de 250 mL	Corning	431096
Sistema de filtrado de 500 mL	Corning	431097
Sistema para qPCR QuantStudio 5	Applied Biosystems	A33621
Sonicador digital	Branson	Sonifier
Termobloque	Thermo Scientific	88870008
Termociclador SimpliAmp	Applied Biosystems	A24811
Transiluminador	Bio-Rad	Gel Doc XR+ Molecular Imager
Tubos 5PRIME Phase Lock Gel Heavy	Quantabio	2302830
Tubos de citometría	Corning	352058
Tubos DNA LoBind	Eppendorf	22431021
Tubos Phasemaker	Invitrogen	A33248
Tubos Protein LoBind	Eppendorf	320108132

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.4 TABLA SUPLEMENTARIA 4. OLIGONUCLEÓTIDOS

Guías de ARN para CRISPR/Cas9		
Gen	**Secuencia directa 5'→3'**	**Secuencia reversa 5'→3'**
Ccm1	CACCGAGGACTTAGCTCTATACTCC	AAACGGAGTATAGAGCTAAGTCCTC
Ccm2	CACCGATGCAGGATTTCCGTCCTAC	AAACGTAGGACGGAAATCCTGCATC
Ccm3	CACCGAGCCGCTTTCATTAAGGTG	AAACCACCTTAATGAAAGCGGCTC
Cebadores para genotipado por PCR		
Gen	**Secuencia directa 5'→3'**	**Secuencia reversa 5'→3'**
Ccm1	TGCTTGCCTTTTAGGGATTG TCCCGGGAGTATAGAGCTAA	CACCAAATTAGGAAAGGCACA
Ccm2	ACAGAGGGAAACCCCATCTT GTAGGACGGAAATCCTGCAT	GCCAAGAACACCCACAAAGT
Ccm3	TCGTAAATCTTACTTTCGGGTAAA GAGAGCCGCTTTCATTAAGGT	CACCATTTGCTGTTGTTTGC
Secuencia de los shARN		
Gen	**Secuencia directa 5'→3'**	
Ccm2	CCGGGACATGAAGGACAGTTACGATCTCGAGATCGTAACTGTCCTTCATGTCTTTTTG	
	AATTCAAAAAGACATGAAGGACAGTTACGATCTCGAGATCGTAACTGTCCTTCATGTC	
Luciferasa	CCGGCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGGTTTTTG	
	AATTCAAAAACCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG	
Cebadores para la amplificación del CDS		
Gen	**Secuencia directa 5'→3'**	**Secuencia reversa 5'→3'**
Ccm2	gctagcGAGGAGGAGGGCAAGAAGG	gctagcGAGGAGGAGGGCAAGAAGG
Cebadores para expresión génica por qPCR		
Gen	**Secuencia directa 5'→3'**	**Secuencia reversa 5'→3'**
Actb	ACCAACTGGGACGACATGGAGAAG	TACGACCAGAGGCATACAGGGACA
Afp	CAGCAGCCTGAGAGTCCATA	GGCGATGGGTGTTTAGAAAG
Ankrd44	GCAGTGCTCAAACTCTGCGA	GTGCGTTTACGTCTTCAGTCT
Apba2	CCAGGTCCTATCCCCCATGA	GTGGGTACATAGCTCTCCAGA
Apoe	CTGACAGGATGCCTAGCCG	CGCAGGTAATCCCAGAAGC
Btbd11	CCCCAGGGTGATATGAACTCT	AGCAGTTTGCGGAACACATTC
Ccm1	AACATCGAAGATGCTTACGTTGC	TCTGTCCTTCAATGGGAACTTCA
Ccm2	ATCCATTCCCGGCTACCTGAA	GCCAACTTGACATTGTAGGCAG
Ccm3	TCACCGAGTCCCTCCTTCG	GCCCGTGCCTTTTCATTTAGG
Cxcl12	TTTCAGATGCTTGACGTTGG	GCGCTCTGCATCAGTGAC
Dnmt1	CGGCTCAAAGACTTGGAAAG	TAGCCAGGTAGCCTTCCTCA
Dnmt3a	GACTCGCGTGCAATAACCTTAG	GGTCACTTTCCCTCACTCTGG
Dnmt3b	CTCGCAAGGTGTGGGCTTTTGTAAC	CTGGGCATCTGTCATCTTTGCACC
Dnmt3l	CCAGGGCAGATTTCTTCCTAAGGTC	TGAGCTGCACAGAGGCATCC
Dusp6	ATAGATACGCTCAGACCCGTG	ATCAGCAGAAGCCGTTCGTT
Dusp9	AGAGTCTGAGTCGGTCATGC	ACTGCTACCCTGGTCGTACA
Eomes	CCAAAACACGGATATCACCC	GACCTCCAGGGACAATCTGA
Epha4	TGGAATTTGCGACGCTGTCA	CACTTCCTCCCACCCTCCTT

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Fgf15	ATGGCGAGAAAGTGGAACGG	CTGACACAGACTGGGATTGCT
Gata4	CCCTACCCAGCCTACATGG	ACATATCGAGATTGGGGTGTCT
Gata6	TTGCTCCGGTAACAGCAGTG	GTGGTCGCTTGTGTAGAAGGA
Gm39469	TGCTGGAGACCTACAAGAACC	TGCCTTTGAGGAAAACTGTG
Ldhb-ps	GAAAAACCTCTCCCGGATTC	CCTCATCGTCCTTCAGCTTC
Macroh2a1	CGGTGGTGAAGTAGGAAACAC	GCTGCCAATGGATGGGAAG
Mest	GTGGTGGGTCCAAGTAGGG	AAGCACAACTATCTCAGGGCT
Mixl1	ATCCGCCCGGACCCTCCAAA	TCGGTTCTGGAACCACACCTGGA
Nanog	TTGCTCTTTCTGTGGGAAGG	CCAGGAAGACCCACACTCAT
Nestin	GAGTTCTCAGCCTCCAGCAG	AGATCGCTCAGATCCTGGAA
Oct4	TCTTTCCACCAGGCCCCCGGCTC	TGCGGGCGGACATGGGGAGATCC
Otx2	GGCCTCACTTTGTTCTGACC	AACTTGCCAGAATCCAGGGT
Pax3	GCAGCGCAGGAGCAGAACCA	GCACTCGGGCCTCGGTAAGC
aRaf	CCTCCTGTTAGTGGGGCTGA	GAGTCATAGACACTCATGCCTTC
bRaf	TGATGCGCTGTCTTCGGAAAT	GCCAGGCTCAAAATCAAACACT
cRaf	TGGACTCAAAGATGCGGTGTT	AAAACCCGGATAGTATTGCTTGT
Sh3bgrl	CTTCTGGCTCTACGGCGATTA	CTCATCCACTTCCGATTCTCTTC
Sox1	GTGACATCTGCCCCCATC	GAGGCCAGTCTGGTGTCAG
Sox17	CGCACGGAATTCGAACAGTA	GTCAAATGTCGGGGTAGTTG
T	GCTTCAAGGAGCTAACTAACGAG	CCAGCAAGAAAGAGTACATGGC
Tet1	ATTGAGGTGGAGAAGTGGG	GGAGAAGGGTTGGTTTGC
Tet2	GGAGAAGGGTTGGTTTGC	GAACTATACATCACTGTGCGTCA
Tet3	CAGAACCAGGTGACCAATGAGG	TGGGCGTGGGCACAGAAGT
Vegfb	GCCAGACAGGGTTGCCATAC	GGAGTGGGATGGATGATGTCAG
Cebadores para expresión génica por PCR		
Gen	**Secuencia directa 5'→3'**	**Secuencia reversa 5'→3'**
Flk1	TTTGGCAAATACAACCCTTCAGA	GCAGAAGATACTGTCACCACC
U6	CTCGCTTCGGCAGCACATATACT	ACGCTTCACGAATTTGCGTGTC
Cebadores para secuenciación Sanger		
Gen / Vector	**Secuencia directa 5'→3'**	
Ccm1	TGCTTGCCTTTTAGGGATTG CACCAAATTAGGAAAGGCACA	
Ccm2	ACAGAGGGAAACCCCATCTT GCCAAGAACACCCACAAAGT	
Ccm3	TCGTAAATCTTACTTTCGGGTAAA CACCATTTGCTGTTGTTTGC	
pJET1.2	CGACTCACTATAGGGAGAGCGGC AAGAACATCGATTTTCCATGGCAG	
pLKO.1 hygro	CAAGGCTGTTAGAGAGATAATTGGA	
SpCas9-2A-Puro V2.0	GAGGGCCTATTTCCCATGAT	

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.5 TABLA SUPLEMENTARIA 5. ANTICUERPOS

Anticuerpo	Aplicación / Concentración	Referencia
IgG	IP: 2 µg/mg proteína	Millipore, 12-371
ACTB monoclonal de ratón	WB: 1/2.000	Santa Cruz Biotechnology, sc-47778
CCM1 policlonal de conejo	WB: 1/2.000	Abcam, ab196025
CCM3 monoclonal de ratón	WB: 1/2.000	Santa Cruz Biotechnology, sc-365586
DNMT3L policlonal de cabra	WB: 1/1.000	Santa Cruz Biotechnology, sc-103480
ERK1/2 monoclonal de ratón	WB: 1/1.000	Santa Cruz Biotechnology, sc-514302
ERK1/2 Thr202/Tyr204; Thr203/Tyr205 monoclonal de ratón	WB: 1/1.000	Santa Cruz Biotechnology, sc-81492
ERK5 policlonal de conejo	WB: 1/1.000	Cell Signaling, 3372
ERK5 Thr218/Tyr220 policlonal de conejo	WB: 1/1.000	Cell Signaling, 3371
EZH2 monoclonal de conejo	WB: 1/1.000	Cell Signaling, 5246S
FLAG-M2 monoclonal de ratón	WB: 1/1.000 ChIP: 1 µg/10 µg ADN IP: 2 µg/mg proteína	Sigma-Aldrich, F1804
GAPDH monoclonal de ratón	WB: 1/10.000	ProteinTech, 10494-1-AP
H2A policlonal de conejo	WB: 1/10.000	GeneTex, GTX129418-25
H2AK119Ub monoclonal de conejo	WB: 1/5.000	Cell Signaling, 8240S
H3 policlonal de conejo	WB: 1/40.000	Abcam, ab1791
H3K27ac monoclonal de conejo	WB: 1/10.000	Abcam, ab218055
H3K27ac monoclonal de conejo	ChIP: 1 µg/5 µg ADN	Abcam, ab4729
H3K27me3 policlonal de conejo	WB: 1/1.000 ChIP: 1 µg/5 µg ADN	Millipore, 07-449
H3K4me1 policlonal de conejo	WB: 1/1.000 ChIP: 1 µg/5 µg ADN	Abcam, ab8895
H3K4me3 policlonal de conejo	WB: 1/1.000 ChIP: 1 µg/5 µg ADN	Abcam, ab8580

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

H3K9ac monoclonal de conejo	WB: 1/2.000 ChIP: 1 µg/5 µg ADN	Abcam, ab218055
H3K9me3 monoclonal de ratón	WB: 1/1.000	Santa Cruz Biotechnology, sc-130356
H3K9me3 monoclonal de ratón	ChIP: 1 µg/5 µg ADN	Abcam, ab8898
HDAC2 policlonal de conejo	WB: 1/2.500	Bethyl, A300-705A-1
JNK monoclonal de ratón	WB: 1/1.000	Santa Cruz Biotechnology, sc-7345
JNK Thr183/Tyr185 monoclonal de ratón	WB: 1/1.000	Santa Cruz Biotechnology, sc-6254
OCT4 monoclonal de ratón	WB: 1/2.000 IF: 1/1.500	Santa Cruz Biotechnology, sc-5279
p38 policlonal de conejo	WB: 1/1.000	Santa Cruz Biotechnology, sc-9212
p38 Thr180/Tyr182 policlonal de conejo	WB: 1/1.000	Santa Cruz Biotechnology, sc-178521-R
SSEA-1 DyLight 488 monoclonal de ratón	IF: 1/1.500	Invitrogen, MA1-022-D488
SUZ12 monoclonal de ratón	WB: 1/1.000	Santa Cruz Biotechnology, sc-271325
TET1 policlonal de conejo	WB: 1/4.000	Millipore, 09-872
TET2 policlonal de conejo	WB: 1/2.000	Abcam, ab124297
ZFP281 monoclonal de ratón	WB: 1/5.000 IF: 1/1.000	Santa Cruz Biotechnology, sc-166933
β-TUBULINA monoclonal de ratón	WB: 1/800	Santa Cruz Biotechnology, sc-55529
CABRA HRP policlonal de asno	WB: 1/5.000	Santa Cruz Biotechnology, sc-2020
CONEJO HRP policlonal de asno	WB: 1/5.000	Invitrogen, A16023
CONEJO TRUEBLOT	WB: 1/5.000	Rockland, 18-8816-33
RATÓN HRP policlonal de asno	WB: 1/5.000	Invitrogen, A16011

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

RATÓN TRUEBLOT	WB: 1/5.000	Rockland, 18-8814-31
RATÓN-RRX policlonal de asno	IF: 1/500	Jackson Immuno Research, 715-295-151

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.6 TABLA SUPLEMENTARIA 6. DATASETS

Descripción	Número acceso	Referencia
SL-PD RNA-seq	GSE97954	Choi *y cols.*, 2017[283]
SL-2i RNA-seq	GSE98517	Hackett *y cols.*, 2017[284]
N2B27-2iL RNA-seq	GSE97954 GSE98517	Choi *y cols.*, 2017[283] Hackett *y cols.*, 2017286
N2B27-2i RNA-seq	GSE98517	Hackett *y cols.*, 2017[284]
N2B27-PD-L RNA-seq	GSE98517	Hackett *y cols.*, 2017[284]
N2B27-CHIR-L RNA-seq	GSE98517	Hackett *y cols.*, 2017[284]
Cribado CRISPR-KO	GSE107060	Li *y cols.*, 2018[297]
Desarrollo temprano RNA-seq	PRJEB7393	Boroviack *y cols.*, 2015[275]
Desarrollo temprano scRNA-seq	GSE22182	Tang *y cols.*, 2011[270]
Diferenciación a EBs micro-array	GSE3749	Hailesellasse Sene *y cols.*, 2007[299]
Diferenciación con RA micro-array	GSE4679	Ivanova *y cols.*,[298]
DNMT3B ChIP-seq (ESCs)	GSE57413	Baubec *y cols.*, 2015[291]
Dnmt3b KO ESCs RNA-seq	GSE72855	Neri *y cols.*, 2017[294]
ECs de cerebro RNA-seq	GSE141980	Marchetti *y cols.*, 2022[305]
ECs de hígado RNA-seq	GSE164006	Furuta *y cols.*, 2021[303]
ECs de pulmón RNA-seq	GSE169532	Rahman *y cols.*, 2021[306]
EPCs RNA-seq	GSE140091	Harland *y cols.*, 2021[304]
ERK2 ChIP-seq (ESCs)	SPR028688	Tee *y cols.*, 2014[105]
Ezh2 KO ESCs RNA-seq	GSE58414	Das *y cols.*, 2015[287]
H3K4me3 ChIP-seq (ESCs)	GSE99530	Mas *y cols.*, 2018[111]
H3K27me3 ChIP-seq (ESCs)	GSE99530	Mas *y cols.*, 2018[111]
NANOG ChIP-seq (ESCs)	GSE44286	Whyte *y cols.*, 2013[276]
OCT4 ChIP-seq (ESCs)	GSE44286	Whyte *y cols.*, 2013[276]
SOX2 ChIP-seq (ESCs)	GSE44286	Whyte *y cols.*, 2013[276]
SUZ12 ChIP-seq (ESCs)	GSE99530	Mas *y cols.*, 2018[111]
TET1 ChIP-seq (ESCs)	GSE104067	Arab *y cols.*, 2019[292]
Tet1 KO ESCs RNA-seq	GSE119443	Reimer Jr *y cols.*, 2019[295]
TET2 ChIP-seq (ESCs)	GSE115972	Rasmussen *y cols.*, 2019[293]
Tet2 KO ESCs RNA-seq	GSE119443	Reimer Jr *y cols.*, 2019[295]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.7 TABLA SUPLEMENTARIA 7. ONTOLOGÍAS

Procesos biológicos	
Descripción	**Código**
Adhesión célula-célula	GO:0098609
Adhesión celular	GO:0007155
Angiogénesis	GO:0001525
Angiogénesis por brotes	GO:0002040
Cascada MAPK	GO:0000165
Compromiso de destino celular	GO:0045165
Desarrollo de la placenta	GO:0001890
Desarrollo ECs	GO:0001885
Desarrollo embrionario en útero	GO:0001701
Desarrollo sistema faríngeo	GO:0060037
Desarrollo vasos sanguíneos	GO:0001568
Diferenciación celular	GO:0030154
Diferenciación celular endotelial	GO:0045446
Diferenciación células troncales	GO:0048863
Fibras de estrés	GO:0001725
Fosforilación	GO:0016310
Fosforilación proteica	GO:0006468
Metilación 5-mC	GO:0090116
Metilación de ADN	GO:0006306
Metilación de histonas	GO:0016571
Metilación H3-H4	GO:0051568
Modificación de histonas	GO:0016570
Morfogénesis tubular	GO:0035239
Morfogénesis vasos sanguíneos	GO:0048514
Regulación de la vía ERK1/2	GO:0070372
Regulación de la vía de ERK5	GO:0070376
Regulación de la vía JNK	GO:0046328
Regulación de la vía p38	GO:0038066
Regulación metilación de ADN	GO:0044030
Regulación negativa proliferación	GO:0008285
Regulación positiva de la vía VEGFR	GO:0030949
Reguladores negativos de ERK1/2	GO:0070373
Reguladores positivos de ERK1/2	GO:0070374
Respuesta a estrés del RE	GO:0034976
Respuesta a estrés oxidativo	GO:0006979
Señalización célula-célula	GO:0007267
Vía de WNT	GO:0016055
Vía FGFR	GO:0008543
Vía NOTCH	GO:0007219
Vía PI3K	GO:0014065
Vía TGFBR	GO:0007179

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Vía TLR3	GO:0034138
Vía TOR	GO:0031929
Vía VEGFR	GO:0048010
Vías de señalización KEGG	
Descripción	**Código**
Adhesión celular	mmu04514
CAMs	mmu04514
Formación eje dorsoventral	mmu04320
Glucólisis / Gluconeogénesis	mmu00010
Regulación de ESCs	mmu04550
Resistencia inhibición TK de EGFR	mmu01521
Uniones estrechas	mmu04530
Vía de ErbB	mmu04012
Vía de FoxO	mmu04068
Vía de HIF-1	mmu04066
Vía de las MAPK	mmu04010
Vía de las pentosas fosfato	mmu00030
Vía de mTOR	mmu04150
Vía de Rap1	mmu04015
Vía de Ras	mmu04014
Vía de TGFβ	mmu04350
Vía de WNT	mmu04310
Funciones moleculares	
Descripción	**Código**
Actividad molecular estructural	GO:0005198
Unión factores crecimiento	GO:0019838
Unión proteínas	GO:0005515
Unión receptores	GO:0005102
Componentes celulares	
Descripción	**Código**
Adhesión focal	GO:0005925
Membrana	GO:0016020
Superficie celular	GO:0009986
Unión celular	GO:0030054
Fenotipos	
Descripción	**Código**
Angiogénesis anormal	MP:0000260
Desarrollo anormal ECs vasculares	MP:0003542
Desarrollo vascular anormal	MP:0000259
Letalidad embrionaria	MP:0008762
Morfología anormal ECs vasculares	MP:0006055
Morfología anormal vasos sanguíneos	MP:0001614

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.8 TABLA SUPLEMENTARIA 8. LISTA DE INTERACTORES DE CCM2

Ranquin	Interactor	Ratio péptidos CCM2/IgG	Ratio PSM CCM2/IgG
1	CCM1	33,00	99,00
2	HIST1H1D	8,00	47,00
3	KRT27	5,00	29,00
4	KRT85	6,00	21,00
5	STK38	13,00	18,00
6	ITGB1BP1	9,00	18,00
7	D1PAS1	8,00	14,00
8	NCOR1	5,00	11,00
9	SH3D19	7,00	10,00
10	PFKFB2	2,00	9,00
11	NELFA	3,00	8,00
12	PDZD11	4,00	7,00
13	BAIAP2	3,67	7,00
14	CAMSAP3	3,29	7,00
15	HSP90B1	3,00	7,00
16	TCOF1	2,29	7,00
17	FMO4	2,00	7,00
18	SPAST	5,00	6,00
19	ARHGAP12	3,00	6,00
20	TPM3	3,00	6,00
21	MAP6	2,33	6,00
22	SLAIN2	2,50	5,75
23	ALB	3,00	5,50
24	FN1	3,33	5,33
25	NARS	5,00	5,00
26	CHAMP1	4,00	5,00
27	GIPC1	4,00	5,00
28	RRP1B	4,00	5,00
29	SETDB1	4,00	5,00
30	PPP1R12A	4,00	5,00
31	SARNP	3,00	5,00
32	TP53BP2	3,00	5,00
33	CD2AP	2,50	5,00
34	NACA	2,10	5,00
35	COL1A2	2,00	5,00
36	CSNK1D	2,00	5,00
37	SKT	2,18	4,88
38	CCDC88C	2,50	4,69
39	SMG7	3,50	4,50
40	ADNP	2,50	4,50
41	COL4A1	3,00	4,00
42	MPP7	3,00	4,00

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

43	RPS2	3,00	4,00
44	TMA7	3,00	4,00
45	MED1	3,00	4,00
46	MOB1B	3,00	4,00
47	NCOA6	2,13	4,00
48	PDLIM7	2,00	4,00
49	KIF5C	2,00	4,00
50	UBN2	2,00	4,00
51	NFRKB	2,33	3,67
52	MELK	2,00	3,67
53	MKI67	2,60	3,60
54	CCAR1	2,50	3,50
55	MAST4	2,50	3,50
56	MED25	2,00	3,50
57	SART1	2,67	3,33
58	KDM3A	2,86	3,29
59	FLNB	3,00	3,20
60	SPTBN1	2,08	3,04
61	ELL	3,00	3,00
62	NDUFA7	3,00	3,00
63	SAMD1	3,00	3,00
64	USP36	3,00	3,00
65	NUP188	2,33	3,00
66	COL4A2	2,00	3,00
67	BCAR3	2,00	3,00
68	LSM12	2,00	3,00
69	TAF6	2,00	3,00
70	TJP1	2,00	3,00
71	VASP	2,00	3,00
72	BAP1	2,00	3,00
73	ENKD1	2,00	3,00
74	TGM1	2,00	3,00
75	ACTN4	2,00	3,00
76	AFF1	2,00	3,00
77	ICT1	2,00	3,00
78	RCOR1	2,00	3,00
79	TMOD3	2,00	3,00
80	SPTAN1	2,23	2,94
81	DLG1	2,80	2,80
82	FLNA	2,00	2,55
83	MAP3K7	2,00	2,50
84	TAB2	2,00	2,50
85	TOMM34	2,00	2,50
86	ZC3H14	2,00	2,50
87	CREBBP	2,00	2,25
88	PRRC2A	2,00	2,25

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

89	PPIA	2,00	2,00
90	ELF2	2,00	2,00
91	FABP5	2,00	2,00
92	PAXIP1	2,00	2,00
93	PLOD3	2,00	2,00
94	WIZ	2,00	2,00
95	CBLB	2,00	2,00
96	COL1A1	2,00	2,00
97	D8ERTD738E	2,00	2,00
98	GRN	2,00	2,00
99	LIN7C	2,00	2,00
100	MAST2	2,00	2,00
101	WIPF1	2,00	2,00
102	ZC3H15	2,00	2,00

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.9 TABLA SUPLEMENTARIA 9. LISTA CON LOS PRINCIPALES GENES DESREGULADOS EN *CCMS* KO IDENTIFICADOS POR RNA-SEQ

Reducidos en ESCs *Ccm1* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Lef1*	-1,49	1,22E-02
2	*Zic2*	-1,22	2,21E-05
3	*Tgm3*	-1,21	1,84E-08
4	*A2m*	-1,01	1,97E-03
5	*Zic5*	-0,94	2,18E-05
6	*Aire*	-0,84	2,52E-05
7	*Lgals9*	-0,79	1,13E-03
8	*Ephx2*	-0,78	8,39E-04
9	*Gstp1*	-0,72	2,65E-07
10	*Gpc3*	-0,68	4,04E-03
11	*Mcf2*	-0,66	6,72E-05
12	*Prdm1*	-0,66	2,98E-02
13	*Igfbp2*	-0,60	6,87E-11
Incrementados en ESCs *Ccm1* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Mcts2*	0,57	7,28E-03
2	*Perp*	0,58	9,31E-03
3	*Slc9a3r2*	0,58	1,49E-02
4	*Nkx6-3*	0,61	9,45E-03
5	*Eif4g3*	0,70	4,38E-09
6	*Nnat*	0,73	7,50E-03
7	*Rdm1*	0,76	7,31E-09
8	*AU015836*	0,79	8,15E-08
9	*Abca1*	0,80	4,46E-06
10	*Map2*	0,81	4,67E-03
11	*Crlf1*	0,84	2,15E-02
12	*Tspan6*	0,88	1,25E-02
13	*Tmem144*	0,95	3,79E-02
14	*Gm43041*	1,07	3,87E-04
15	*Stard8*	1,32	2,48E-02
16	*Rhob*	1,38	6,61E-30
17	*Tbx20*	1,54	2,05E-02
18	*Gm11635*	1,63	1,36E-02
19	*Mest*	3,79	2,06E-73
Reducidos en ESCs *Ccm2* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Gm36283*	-5,64	9,62E-27
2	*Il12rb2*	-5,19	1,34E-07
3	*Rbbp8nl*	-4,34	1,73E-07

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4	*Pitx2*	-4,20	3,45E-14
5	*G730013B05Rik*	-4,11	2,34E-05
6	*Prom2*	-3,83	7,49E-05
7	*Gm45752*	-3,29	6,99E-07
8	*Upk2*	-3,18	3,22E-05
9	*Ctxn3*	-3,17	5,77E-04
10	*Bfsp1*	-3,12	8,01E-04
11	*Spink1*	-3,09	9,27E-26
12	*Meis2*	-3,03	3,63E-46
13	*Bnipl*	-2,97	2,39E-37
14	*Tmem40*	-2,96	2,97E-14
15	*Klhl13*	-2,95	3,31E-106
16	*Themis*	-2,93	1,25E-06
17	*Rab25*	-2,82	8,42E-24
18	*2700033N17Rik*	-2,78	5,53E-03
19	*Bnip3*	-2,68	1,17E-27
20	*Ndufa4l2*	-2,65	5,08E-10
21	*Tex13c2*	-2,62	1,31E-04
22	*2610528J11Rik*	-2,58	2,00E-06
23	*Ldhc*	-2,50	4,46E-05
24	*Slc23a1*	-2,49	3,90E-15
25	*Aire*	-2,49	1,43E-54
26	*Gm32885*	-2,47	7,18E-05
27	*Dppa3*	-2,46	2,23E-03
28	*Rnf128*	-2,41	2,19E-12
29	*Clca4a*	-2,40	8,43E-04
30	*Esrp1*	-2,38	1,64E-75
31	*Tmem54*	-2,30	5,95E-06
32	*Sdr42e1*	-2,29	3,17E-04
33	*Hopx*	-2,29	6,35E-66
34	*Tmem30b*	-2,27	5,99E-18
35	*Lef1*	-2,26	1,13E-06
36	*Grin1*	-2,24	1,48E-11
37	*Gm20005*	-2,21	1,10E-02
38	*AU015836*	-2,18	3,88E-37
39	*Susd1*	-2,16	1,79E-17
40	*Tmem154*	-2,15	2,22E-03
41	*Cdkn2b*	-2,13	3,85E-05
42	*1700019A02Rik*	-2,11	1,77E-04
43	*Vrtn*	-2,10	4,04E-04
44	*Prss8*	-2,08	6,71E-14
45	*Myzap*	-2,07	8,60E-03
46	*Gm7765*	-2,01	1,69E-03
47	*Hipk4*	-2,00	3,88E-06
48	*Shroom4*	-1,99	2,61E-11
49	*Gm16136*	-1,99	2,21E-03

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

50	*Gm44771*	-1,98	1,48E-02
Incrementados en ESCs *Ccm2* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Atp1a3*	2,13	1,82E-109
2	*Klhdc7a*	2,15	9,89E-04
3	*Zfp711*	2,16	4,39E-03
4	*Tns1*	2,18	5,95E-84
5	*Oaf*	2,19	1,09E-05
6	*Cxcl14*	2,22	1,30E-02
7	*Hoxc12*	2,25	2,08E-03
8	*Myot*	2,25	6,06E-03
9	*Speg*	2,26	1,44E-34
10	*Grm4*	2,26	5,70E-04
11	*Mir155hg*	2,26	2,28E-02
12	*Hdx*	2,29	3,00E-06
13	*Gfra1*	2,29	2,28E-02
14	*Atp2b4*	2,31	3,97E-19
15	*Nlrp4a*	2,36	2,25E-02
16	*1700016D06Rik*	2,38	1,61E-03
17	*Vstm2l*	2,40	9,71E-05
18	*En1*	2,40	2,54E-03
19	*Gm3739*	2,42	1,29E-03
20	*Nol4l*	2,44	2,12E-18
21	*Calb2*	2,45	9,74E-17
22	*Eya1*	2,49	1,36E-24
23	*Zscan4a*	2,50	6,02E-03
24	*Ldhb-ps*	2,51	4,38E-03
25	*Clec2l*	2,52	9,74E-03
26	*Hmgn3*	2,54	1,78E-05
27	*Slit2*	2,55	1,11E-03
28	*Frrs1l*	2,59	1,31E-05
29	*AI854703*	2,66	1,30E-11
30	*Peg12*	2,72	3,48E-05
31	*Dock10*	2,76	1,06E-04
32	*BC034090*	2,80	1,26E-03
33	*Ky*	2,83	2,03E-03
34	*Cacna1c*	2,85	1,75E-02
35	*Kif1a*	2,87	6,34E-14
36	*Zscan4c*	2,93	1,53E-02
37	*Crabp1*	3,00	1,07E-48
38	*Tram1l1*	3,14	5,51E-04
39	*Tgfbi*	3,20	3,47E-59
40	*Zscan4d*	3,25	8,68E-03
41	*Sh3bgrl*	3,32	3,83E-30
42	*Trpa1*	3,63	2,66E-12

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

43	*Asic4*	4,76	3,04E-03
44	*Mest*	5,02	7,67E-203
45	*Gm29650*	5,26	9,33E-04
46	*Uty*	5,32	9,34E-52
47	*Eif2s3y*	5,44	3,00E-212
48	*Wfikkn2*	5,93	8,51E-04
49	*Ddx3y*	6,51	1,20E-17
50	*Uba1y*	6,92	7,04E-05
Reducidos en ESCs *Ccm3* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Eif2s3y*	-3,60	8,94E-06
2	*Tgm3*	-2,17	6,86E-30
3	*Lrrn2*	-2,06	5,97E-10
4	*Gm3333*	-1,71	1,68E-08
5	*Gm7896*	-1,69	2,69E-03
6	*Pou2f3*	-1,63	3,84E-06
7	*Klk1*	-1,63	9,34E-03
8	*Zscan5b*	-1,62	6,56E-03
9	*Spn*	-1,56	6,30E-11
10	*Treml2*	-1,55	1,82E-04
11	*Gpr83*	-1,50	7,25E-03
12	*Foxn4*	-1,43	1,68E-05
13	*Jakmip1*	-1,41	1,57E-02
14	*Zfp951*	-1,37	1,92E-03
15	*Igsf21*	-1,33	5,52E-05
16	*Myo1f*	-1,27	8,48E-11
17	*Gstp3*	-1,26	1,08E-02
18	*Kit*	-1,26	4,61E-19
19	*Fhod3*	-1,23	1,21E-08
20	*Pdyn*	-1,21	2,51E-03
21	*Nt5e*	-1,20	8,58E-03
22	*9330159F19Rik*	-1,20	9,16E-06
23	*A2m*	-1,18	6,96E-08
24	*Tns1*	-1,18	1,51E-08
25	*Prkca*	-1,17	2,59E-03
26	*Pla2g4e*	-1,15	4,10E-03
27	*Col13a1*	-1,15	5,95E-04
28	*Rbfaos*	-1,13	2,62E-05
29	*Gstp2*	-1,09	8,29E-04
30	*Ggt1*	-1,09	1,09E-03
31	*1700097N02Rik*	-1,09	1,45E-03
32	*Vegfc*	-1,06	3,49E-11
33	*Slc6a1*	-1,05	2,48E-09
34	*Atp6v0a4*	-1,04	2,77E-04
35	*Kdelr3*	-1,04	3,90E-06

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

36	*Ephx2*	-1,04	1,85E-09
37	*Ptgs1*	-1,03	8,58E-03
38	*Myo1g*	-1,03	1,31E-04
39	*Itpka*	-1,02	2,19E-04
40	*Hydin*	-1,02	1,68E-02
41	*Sema5b*	-1,01	6,23E-15
42	*Zic2*	-1,00	2,85E-07
43	*Soga3*	-1,00	1,06E-15
44	*Lonrf2*	-1,00	7,69E-03
45	*Axin2*	-0,99	2,01E-09
46	*Tgfbi*	-0,94	2,60E-02
47	*Gm48185*	-0,92	2,01E-02
48	*Dnah8*	-0,91	5,76E-12
49	*Shisa2*	-0,91	2,10E-03
50	*Aire*	-0,89	1,13E-11
Incrementados en ESCs *Ccm3* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Lfng*	1,73	1,24E-03
2	*AU015836*	1,73	3,19E-79
3	*Tbx20*	1,75	1,82E-04
4	*Atp10a*	1,75	2,33E-10
5	*Parva*	1,75	7,94E-06
6	*Kctd12b*	1,76	1,10E-02
7	*Rhob*	1,78	1,69E-82
8	*Tent5a*	1,81	8,03E-03
9	*App*	1,82	2,72E-10
10	*Jun*	1,84	7,57E-17
11	*Gm43041*	1,84	1,95E-21
12	*Slc25a24*	1,84	4,31E-03
13	*Anxa1*	1,85	2,52E-07
14	*Dmrta2*	1,85	4,32E-03
15	*Pax6*	1,89	1,83E-09
16	*Aass*	1,92	3,01E-14
17	*Krt19*	1,92	5,35E-08
18	*Fgfr2*	1,94	3,94E-05
19	*Zfp608*	1,96	6,63E-03
20	*Nes*	1,96	1,90E-18
21	*Meis1*	1,96	6,11E-09
22	*Ccn2*	1,99	1,32E-36
23	*S100a6*	2,08	1,01E-72
24	*Has2*	2,10	1,31E-03
25	*Arntl2*	2,12	1,50E-06
26	*Fgfbp1*	2,12	7,45E-11
27	*Msln*	2,26	3,93E-32
28	*Tmprss5*	2,30	3,45E-07

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

29	*Ccn1*	2,37	2,82E-34
30	*Wnt7b*	2,38	1,81E-05
31	*Hdx*	2,43	7,04E-07
32	*Amot*	2,46	3,77E-11
33	*Kif1a*	2,47	2,73E-09
34	*Ptgs2*	2,52	5,04E-04
35	*Pla2g4a*	2,57	1,88E-06
36	*Wls*	2,57	1,45E-08
37	*Gm28625*	2,76	2,44E-07
38	*Anxa3*	2,78	8,80E-17
39	*Gna14*	2,84	7,63E-18
40	*Stard8*	2,84	2,04E-20
41	*Sh3bgrl*	2,92	4,96E-20
42	*Sema6a*	3,09	1,31E-07
43	*C1qtnf9*	3,15	4,83E-05
44	*Tagln*	3,18	1,76E-25
45	*Adamts1*	3,60	6,10E-23
46	*Fndc3c1*	3,60	3,49E-10
47	*Thbs1*	3,82	2,10E-25
48	*Igf2*	4,57	2,46E-08
49	*Slc39a8*	4,67	4,17E-13
50	*Acta2*	5,66	5,64E-09
Reducidos en EPCs *Ccm1* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Lipc*	-2,26	1,28E-03
2	*Col13a1*	-2,01	2,17E-12
3	*Trim47*	-1,98	7,33E-20
4	*Hal*	-1,98	4,06E-06
5	*Slc6a7*	-1,96	9,38E-13
6	*Igsf21*	-1,93	8,40E-03
7	*Crtac1*	-1,81	3,31E-12
8	*Mt3*	-1,79	8,29E-07
9	*Dusp9*	-1,74	2,41E-22
10	*Grid2ip*	-1,72	1,22E-03
11	*Car8*	-1,69	4,18E-04
12	*Gmnc*	-1,68	4,23E-04
13	*Amdhd1*	-1,57	3,28E-03
14	*Fbxo2*	-1,53	8,26E-08
15	*En2*	-1,53	8,65E-14
16	*Cdx1*	-1,53	5,51E-25
17	*Kank4*	-1,52	4,34E-08
18	*Khdc3*	-1,52	4,81E-09
19	*Tmem151b*	-1,49	1,02E-03
20	*Rtn1*	-1,47	1,95E-15
21	*Inhbb*	-1,45	3,32E-14

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

22	*Shisa8*	-1,44	7,13E-06
23	*Myl7*	-1,44	9,38E-03
24	*Pdgfc*	-1,44	8,78E-21
25	*Mras*	-1,42	3,04E-07
26	*Pttg1*	-1,39	2,58E-30
27	*Nkx6-3*	-1,38	1,24E-07
28	*Tgfbi*	-1,38	4,58E-03
29	*Lama1*	-1,37	1,60E-07
30	*Gm34299*	-1,35	2,47E-02
31	*Tgfb1*	-1,35	4,38E-07
32	*Sema3f*	-1,34	2,85E-06
33	*Bcl11b*	-1,34	5,02E-08
34	*Krt42*	-1,33	7,20E-05
35	*Atp1a3*	-1,33	1,84E-23
36	*Gm53*	-1,33	1,03E-07
37	*Fbxo44*	-1,28	1,15E-04
38	*Itpka*	-1,27	4,17E-08
39	*Pde4c*	-1,25	2,37E-06
40	*Cpne5*	-1,25	1,18E-16
41	*Cd37*	-1,25	2,38E-18
42	*Gstp2*	-1,24	6,43E-13
43	*Tmem181b-ps*	-1,22	2,68E-24
44	*Hck*	-1,22	8,40E-14
45	*Slc16a9*	-1,21	3,93E-06
46	*Extl1*	-1,21	1,69E-02
47	*Stambpl1*	-1,21	8,02E-04
48	*Camk2n1*	-1,20	5,65E-08
49	*Sfmbt2*	-1,19	7,00E-80
50	*Nxph3*	-1,18	9,15E-03
Incrementados en EPCs *Ccm1* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Plac8*	1,82	4,23E-05
2	*Efna5*	1,83	4,37E-04
3	*Gbp3*	1,84	1,16E-03
4	*Wnt6*	1,84	2,72E-04
5	*Bcl11a*	1,85	1,50E-10
6	*Oasl2*	1,86	6,79E-10
7	*Atf3*	1,86	3,27E-07
8	*Tgfb1i1*	1,88	3,94E-10
9	*Stbd1*	1,88	1,74E-05
10	*Ctso*	1,91	1,22E-02
11	*App*	1,93	1,54E-27
12	*Ptprt*	1,94	2,09E-12
13	*Stra6*	1,95	1,71E-03
14	*Serpinb6c*	1,96	8,82E-04

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

15	*Ebf1*	1,96	1,12E-26
16	*Has2*	1,97	3,33E-20
17	*Sema6a*	2,03	2,90E-23
18	*Aldh1a2*	2,03	1,85E-03
19	*Lrrc18*	2,04	5,31E-03
20	*Nlrc5*	2,07	1,31E-05
21	*Dppa3*	2,09	1,25E-09
22	*Fosl2*	2,10	1,46E-04
23	*1700019A02Rik*	2,12	3,25E-03
24	*Nkx6-2*	2,16	2,75E-03
25	*Adamts6*	2,18	3,14E-14
26	*Psmb8*	2,20	1,42E-08
27	*Nxf3*	2,24	8,29E-09
28	*Krt19*	2,32	2,60E-10
29	*Fmr1nb*	2,33	8,62E-13
30	*Htra1*	2,45	3,65E-03
31	*Cdh13*	2,48	1,43E-05
32	*Ltb*	2,57	1,66E-04
33	*Ccn1*	2,58	8,06E-14
34	*Gad1*	2,61	5,48E-05
35	*Pdlim3*	2,63	8,35E-09
36	*1700013H16Rik*	2,70	3,03E-07
37	*Acta1*	2,75	2,54E-04
38	*Ifih1*	2,78	1,09E-03
39	*Irgm2*	2,80	5,02E-08
40	*Gm43041*	3,00	1,88E-23
41	*Cnn1*	3,12	2,41E-10
42	*Gbp2*	3,15	1,71E-38
43	*Thbs1*	3,16	2,17E-17
44	*Myl9*	3,19	2,13E-26
45	*Gbp2b*	3,26	6,12E-25
46	*Tagln*	3,51	6,39E-21
47	*Acta2*	3,88	1,46E-04
48	*Mest*	5,36	1,66E-130
49	*Actc1*	5,85	3,65E-04
50	*Ppargc1a*	8,60	7,34E-07
Reducidos en EPCs *Ccm2* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Gm8574*	-8,00	3,29E-07
2	*Gm36283*	-6,60	5,28E-24
3	*Il12rb2*	-6,46	1,83E-36
4	*Fcgr3*	-6,00	1,49E-04
5	*Ndufa4l2*	-4,78	2,34E-20
6	*Nkx6-3*	-4,32	2,47E-08
7	*Grin1*	-4,25	1,15E-83

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8	*Sgcz*	-4,22	1,32E-03
9	*Hapln1*	-4,00	1,62E-03
10	*Gm9165*	-3,91	3,10E-03
11	*Jaml*	-3,71	3,92E-03
12	*Aire*	-3,65	2,37E-63
13	*Gpx2*	-3,63	2,72E-50
14	*Mir669o*	-3,17	1,48E-03
15	*Nxph4*	-3,07	2,79E-06
16	*Foxq1*	-3,06	1,30E-03
17	*Tbx15*	-3,00	2,44E-18
18	*Mir669i*	-2,91	8,00E-08
19	*mmu-mir-467b*	-2,88	2,49E-03
20	*Grin1os*	-2,78	2,78E-05
21	*Pttg1*	-2,76	1,50E-61
22	*Ly6a*	-2,73	7,33E-07
23	*Dkk1*	-2,69	3,79E-05
24	*Khdc3*	-2,68	1,70E-27
25	*Rpl39l*	-2,67	3,54E-24
26	*En2*	-2,64	9,70E-53
27	*Mir466b-8*	-2,61	6,89E-03
28	*Mir669a-2*	-2,55	1,21E-02
29	*Camk2n1*	-2,47	8,34E-33
30	*AU015836*	-2,46	6,31E-12
31	*Gstp2*	-2,46	2,33E-35
32	*Nnat*	-2,44	4,19E-20
33	*Slc52a3*	-2,41	2,62E-14
34	*C130073F10Rik*	-2,41	1,53E-04
35	*Mocos*	-2,38	2,33E-23
36	*Gfod1*	-2,38	2,21E-28
37	*mmu-mir-467a-2*	-2,36	6,83E-03
38	*Ankrd34a*	-2,33	1,91E-04
39	*Gm20610*	-2,32	5,41E-05
40	*Ak4*	-2,30	7,39E-229
41	*Clca3b*	-2,24	6,08E-06
42	*Sfmbt2*	-2,23	1,74E-23
43	*Mir466h*	-2,17	1,91E-02
44	*Trim47*	-2,16	2,19E-32
45	*Scnn1a*	-2,15	1,36E-07
46	*Gstp-ps*	-2,14	2,02E-03
47	*Aldh3b2*	-2,11	6,05E-03
48	*Gm29456*	-2,10	1,41E-02
49	*Gm37827*	-2,08	2,95E-02
50	*Lipc*	-2,08	1,73E-05
Incrementados en EPCs *Ccm2* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1	*Gm3558*	2,73	3,33E-06
2	*Gbp2b*	2,75	1,45E-12
3	*Gad1*	2,79	9,63E-10
4	*Gm12198*	2,79	5,08E-03
5	*Kif1a*	2,79	8,57E-115
6	*Foxp2*	2,81	8,34E-04
7	*Slc16a2*	2,82	7,22E-17
8	*Trank1*	2,82	5,80E-08
9	*Tlr5*	2,83	3,91E-04
10	*Zfhx4*	2,85	4,65E-13
11	*Wnt6*	2,87	2,82E-10
12	*Gm3667*	2,89	1,45E-04
13	*Thbs1*	2,93	5,34E-10
14	*Gm37728*	2,98	8,12E-04
15	*Ptchd1*	2,99	4,48E-03
16	*Tagln*	3,00	1,58E-09
17	*Gm50072*	3,06	1,72E-03
18	*Gm3411*	3,07	4,05E-06
19	*2310002F09Rik*	3,10	5,84E-03
20	*Loxl4*	3,10	1,54E-04
21	*Gm43041*	3,12	1,08E-50
22	*Plagl1*	3,15	8,96E-09
23	*Gm43909*	3,16	8,23E-04
24	*Klk8*	3,17	8,01E-10
25	*Bhlhe41*	3,17	5,60E-25
26	*Nrsn1*	3,18	8,69E-13
27	*Gm3264*	3,19	2,13E-03
28	*Dach1*	3,21	6,80E-04
29	*Gm44758*	3,22	5,07E-03
30	*Irs1*	3,40	8,21E-42
31	*Npr3*	3,57	5,28E-12
32	*Gm50094*	3,65	1,87E-03
33	*A330076H08Rik*	3,65	3,31E-04
34	*Snap25*	3,69	2,38E-13
35	*Sox5*	3,84	1,69E-06
36	*Lonrf3*	3,97	4,01E-13
37	*Pou3f3*	3,99	2,10E-03
38	*Gm9801*	4,11	2,33E-13
39	*Piwil4*	4,27	5,47E-13
40	*Capns2*	4,52	4,93E-04
41	*Xlr3a*	4,76	1,23E-09
42	*Gm10406*	5,06	1,49E-06
43	*Actc1*	5,24	3,32E-04
44	*Gm29650*	6,38	1,59E-04
45	*A230057D06Rik*	6,89	2,15E-06
46	*Mest*	7,53	1,54E-22

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

47	*Uba1y*	8,05	1,41E-08
48	*Ddx3y*	12,50	1,25E-21
49	*Uty*	12,72	8,49E-24
50	*Eif2s3y*	14,48	2,26E-30
Reducidos en EPCs *Ccm3* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Gm47896*	-6,05	6,40E-04
2	*Il12rb2*	-5,47	2,46E-23
3	*Gm36283*	-4,99	1,61E-22
4	*Inhbb*	-4,93	2,12E-65
5	*Nkx6-3*	-4,59	1,07E-32
6	*Olfr915*	-4,56	6,89E-03
7	*Hal*	-4,09	6,41E-14
8	*Slc4a9*	-3,94	1,32E-07
9	*Mir124a-1hg*	-3,83	4,30E-27
10	*Gm47903*	-3,68	3,10E-03
11	*Gm34299*	-3,50	4,63E-05
12	*Myo1f*	-3,48	1,26E-10
13	*Gm14221*	-3,44	2,29E-12
14	*Trim47*	-3,23	3,46E-41
15	*Zfyve28*	-3,17	1,17E-28
16	*Kank4*	-3,13	7,15E-23
17	*Omp*	-3,11	7,82E-06
18	*Gjb1*	-3,10	4,56E-13
19	*Lgals9*	-3,05	7,75E-66
20	*Cd37*	-3,04	1,53E-73
21	*Grin1*	-3,04	2,96E-28
22	*Grin1os*	-3,00	7,73E-04
23	*Gm21028*	-2,97	1,10E-07
24	*Lipc*	-2,96	1,55E-05
25	*Crtac1*	-2,93	3,24E-06
26	*Slc6a1*	-2,93	1,67E-12
27	*Tbx15*	-2,92	7,75E-11
28	*Gm20610*	-2,86	7,77E-03
29	*Hhip*	-2,84	3,71E-09
30	*Gfod1*	-2,80	9,72E-20
31	*Ankrd35*	-2,73	3,40E-16
32	*Dpp4*	-2,71	6,91E-90
33	*Gm44997*	-2,70	1,46E-12
34	*Khdc3*	-2,66	1,90E-21
35	*Fam81a*	-2,66	2,03E-53
36	*Gm15513*	-2,63	4,89E-07
37	*Mymx*	-2,62	8,05E-93
38	*Mog*	-2,61	4,80E-04
39	*Pdgfc*	-2,59	1,17E-53

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

40	*Krt42*	-2,58	3,45E-13
41	*A930004D18Rik*	-2,55	3,05E-34
42	*Tbx3*	-2,54	4,48E-117
43	*Tcl1*	-2,53	9,12E-88
44	*Itpka*	-2,52	4,43E-23
45	*AA543401*	-2,50	7,31E-04
46	*Slc6a7*	-2,49	2,21E-19
47	*Ndufa4l2*	-2,48	7,86E-09
48	*Amdhd1*	-2,48	1,88E-06
49	*Tspan1*	-2,48	6,95E-06
50	*Cobl*	-2,46	1,00E-142
Incrementados en EPCs *Ccm3* KO *vs*. WT			
Ranquin	**Gen**	**FC (log2)**	**FDR**
1	*Avpr1b*	6,22	2,77E-08
2	*Gm3264*	6,25	3,25E-04
3	*Trim30a*	6,28	1,11E-21
4	*Sh3rf2*	6,28	3,77E-04
5	*Cd248*	6,32	2,77E-05
6	*Rtp4*	6,32	9,63E-19
7	*Sostdc1*	6,34	2,54E-04
8	*Adh1*	6,35	2,46E-04
9	*Mest*	6,38	3,39E-213
10	*Alx4*	6,39	2,25E-04
11	*4933427D06Rik*	6,42	1,77E-05
12	*Rgs4*	6,56	1,53E-04
13	*Ifi27*	6,56	1,10E-05
14	*Synpo2l*	6,60	1,57E-37
15	*Cxcl1*	6,64	1,16E-04
16	*Xlr4a*	6,64	1,16E-04
17	*Tagln*	6,65	5,32E-36
18	*H2-T10*	6,70	7,08E-06
19	*Xlr3a*	7,03	3,76E-15
20	*Frmpd2*	7,16	2,82E-05
21	*Ptchd1*	7,17	2,56E-05
22	*Isg15*	7,17	7,04E-16
23	*Ltb*	7,18	2,49E-65
24	*Tnni2*	7,24	2,40E-05
25	*Ccn1*	7,26	2,47E-69
26	*Tacstd2*	7,27	1,01E-06
27	*Acta1*	7,30	1,99E-54
28	*Gm29650*	7,45	1,06E-05
29	*Usp26*	7,59	1,17E-28
30	*Shh*	7,62	6,39E-06
31	*Gbp11*	7,63	6,37E-06
32	*Mat1a*	7,65	5,82E-06

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

33	*Tnc*	7,66	2,00E-07
34	*Cxcl10*	7,70	1,19E-12
35	*Ankrd1*	7,82	3,72E-06
36	*Tnf*	7,88	2,73E-06
37	*Pappa*	8,02	1,66E-06
38	*Cxcl11*	8,11	1,36E-06
39	*H2-M2*	8,17	1,10E-06
40	*Acta2*	8,22	2,84E-26
41	*Msln*	8,26	8,93E-07
42	*Nog*	8,38	1,14E-08
43	*Actc1*	9,08	4,48E-10
44	*Zfy1*	9,20	1,93E-08
45	*Wnt10b*	9,73	2,65E-09
46	*Kdm5d*	10,38	1,48E-10
47	*Uba1y*	10,55	6,71E-11
48	*Uty*	12,93	5,66E-16
49	*Ddx3y*	13,87	3,51E-18
50	*Eif2s3y*	14,56	5,25E-20

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9. LISTA DE ABREVIATURAS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2iL: medio de cultivo *naïve*

5-caC: 5-carboxilcitosina

5-fC: 5-formilcitosina

5-hmC: 5-hidroximetilcitosina

5-mC: 5-metilcitosina

ADNc: ADN complementario

AP: fosfatasa alcalina (del inglés *alkaline phosphatase*)

ARNr: ARN ribosómico

BMP: proteína morfogénica ósea (del inglés *bone morphogenic protein*)

bp: pares de bases

BSA: albúmina de suero bovino (del inglés *bovine serum albumin*)

CCM: malformaciones cavernosas cerebrales (del inglés *cerebral cavernous malformations*)

CDS: secuencia codificante

cél.: célula

ChEA: análisis de enriquecimiento de factores de transcripción (del inglés *chromatin enrichment analysis*)

ChIP-seq: inmunoprecipitación de cromatina (ChIP) seguida de secuenciación

DAPI: 4’,6-diamidino-2-fenilindol

DAVID: del inglés *database for annotation, visualization, and integrated discovery*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

DEGs: genes diferencialmente expresados (del inglés *differential expressed genes*)

DMEM: medio Eagle modificado de Dulbecco

DMSO: dimetilsulfóxido

DNMT: ADN metiltransferasas (del inglés *de novo DNA methyltransferase*)

E. pr.: endodermo primitivo

EBs: cuerpos embrionarios (del inglés *embryoid bodies*)

ECs: células endoteliales (del inglés *endothelial cells*)

ELISA: ensayo de inmunoadsorción ligado a enzima

EPCs: células progenitoras endoteliales (del inglés *endothelial progenitor cells*)

Epi.: epiblasto

Epi. tar.: epiblasto tardío

Epi. tem.: epiblasto temprano

EpiSCs: células troncales de epiblasto (del inglés *Epiblast Stem Cells*)

ERK: quinasa regulada por señal extracelular ERK (del inglés *extracelular signal regulated kinase*)

ESCs: células troncales embrionarias (del inglés *embryonic stem cells*)

FA: medio de cultivo *primed*

FACS: citometría de flujo activada por fluorescencia

FBS: suero fetal bovino

FC: valor de cambio (del inglés *fold-change*)

FDR: tasa de descubrimientos falsos (del inglés *false discovery rate*)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FGF: factor de crecimiento fibroblástico (del inglés *fibroblast growth factor*)

GFP: proteína fluorescente verde (del inglés *green fluorescent protein*)

GO: ontología génica (del inglés *gene ontology*)

GSEA: análisis de enriquecimiento de conjuntos de genes (del inglés *gene set enrichment analysis*)

GSK3β: glucógeno sintasa quinasa-3 beta

H2AK119ub: grupo ubiquitina en la lisina 119 de la histona H2A

H3K27ac: grupo acetilo en la lisina 27 de la histona H3

H3K27me3: grupo metilo en la lisina 27 de la histona H3

H3K4me3: tres grupos metilo en la lisina 4 de la histona H3

H3K4me1: grupo metilo en la lisina 4 de la histona H3

H3K9ac: grupo acetilo en la lisina 9 de la histona H3

H3K9me3: grupo metilo en la lisina 9 de la histona H3

HOMER: del inglés *hypergeometric optimization of motif enrichment*

ICM: masa celular interna (del inglés *inner cell mass*)

IP: inmunoprecipitación

KEGG: vías de señalización KEGG (del inglés *kyoto encyclopedia of genes and genomes*)

KO: *knock-out*

LB: medio Luria-Bertani

LIF: factor inhibidor de la leucemia

M: mórula

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

MAPK: proteínas quinasa activadas por mitógenos (del inglés *mitogen-activated protein kinases*)

MEK: quinasa proteína-quinasa activada por mitógenos MEK (del inglés *mitogen-activated protein kinase kinase*)

MLL: metiltransferasas de histonas (del inglés *mixed lineage leukemia*)

MPO: ontología de fenotipos de mamíferos (del inglés *mammalian phenotype ontology*)

NES: valor de enriquecimiento normalizado (del inglés *normalized enrichment score*)

nLC-MS/MS: análisis de espectrometría de masas en tándem por cromatografía de nanolíquidos

PBS: tampón fosfato salino

PCA: análisis de componentes principales (del inglés *principal component analysis*)

PGCs: células germinales primordiales (del inglés *primordial germinal cells*)

PCR: reacción en cadena de la polimerasa

PCV: volumen de células sedimentadas (del inglés *packed cell volume*)

PEI: polietilenimina

pept.: péptidos

PIC: cocktail inhibidor de proteasas

PNK: polinucleótido quinasa

PRC: complejo represor de *Polycomb* (del inglés *Polycomb repressor complex*)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PSM: número total de secuencias peptídicas identificadas (del inglés *peptide spectrum matches*)

qPCR: reacción en cadena de la polimerasa cuantitativa

RA: ácido retinoico (del inglés *retinoic acid*)

RNA-seq: secuenciación masiva de ARN

rpm: revoluciones por minuto

RT: temperatura ambiente (del inglés *room temperature*)

s.e.m.: error estándar de la media (del inglés *standard error of the mean*)

shARN: ARN de horquilla corta

SL: medio de cultivo estándar pluripotente (conteniendo suero y LIF)

TE: trofectodermo

TET: enzimas metilcitosina dioxigenasas (del inglés *ten-eleven translocation*)

TSS: sitio de inicio de la transcripción

WB: western blot

WT: *wildtype*

Z: zigoto

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10. LISTA DE FIGURAS Y TABLAS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figura I1. Desarrollo embrionario temprano.

Figura I2. Derivación *ex vivo* de células pluripotentes.

Figura I3. Circuito autorregulatorio de los factores centrales en el control transcripcional del estado pluripotente.

Figura I4. Ciclo de la metilación y desmetilación activa del ADN en células de mamífero.

Figura I5. Composición de los complejos proteicos *Polycomb*.

Figura I6. Activación de las vías que rigen el espectro pluripotente.

Figura I7. Vía de señalización de LIF.

Figura I8. Vía de señalización de NODAL/ACTIVINA/BMP4.

Figura I9. Vía de señalización de WNT.

Figura I10. Vía de señalización de FGF-ERK1/2.

Figura I11. Diferenciación endotelial a partir de células pluripotentes.

Figura I12. Dominios e interacciones de las proteínas CCM descritas en células endoteliales.

Figura I13. Funciones de las proteínas CCM en células endoteliales.

Figura I14. Relevancia de las proteínas CCM durante el desarrollo embrionario murino.

Figura I15. Modelo de la aparición y crecimiento de malformaciones cavernosas cerebrales.

Tabla I1. Comparación de los distintos estados alternativos pluripotentes.

Figura R1. El análisis de los patrones de expresión de los genes *Ccm* revela su presencia en células pluripotentes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figura R2. Generación de líneas KO para *Ccm1*, *Ccm2* y *Ccm3* en ESCs.

Figura R3. La ausencia de cada gen *Ccm* no conlleva la pérdida de pluripotencia.

Figura R4. La eliminación de las proteínas CCM no afecta a la proliferación de las células pluripotentes.

Figura R5. CCM2 afecta negativamente a la capacidad de autorrenovación de las ESCs.

Figura R6. La eliminación de las proteínas CCM altera el perfil transcripcional de las células pluripotentes.

Figura R7. Identificación de los cambios transcripcionales promovidos por la pérdida de función de cada gen *Ccm* en ESCs.

Figura R8. Análisis comparativo del perfil transcripcional alterado en ausencia de CCM2 y CCM3.

Figura R9. Las ESCs KO para *Ccm2* son transcripcionalmente similares a las células de la masa interna.

Figura R10. CCM2 reprime la expresión de genes asociados con la pluripotencia *naïve*.

Figura R11. La eliminación de CCM2 conlleva un aumento de la población de ESCs con niveles elevados de NANOG.

Figura R12. La ausencia de CCM2 favorece la reprogramación a pluripotencia *naïve* tras diferenciación celular.

Figura R13. La falta de CCM2 dificulta la obtención del estado pluripotente *primed*.

Figura R14. La pérdida de función de CCM2 no altera la proliferación celular durante la reprogramación a estados alternativos pluripotentes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figura R15. CCM2 es necesario para el mantenimiento del estado pluripotente *primed*.

Figura R16. CCM2 se encuentra presente en la cromatina de ESCs.

Figura R17. CCM2 no controla directamente la transcripción de los genes desregulados como resultado de su ausencia en ESCs.

Figura R18. Identificación de procesos biológicos alterados y vías moleculares afectadas en *Ccm2* KO ESCs.

Figura R19. Identificación de procesos biológicos asociados a genes reprimidos y genes activados en ausencia de CCM2.

Figura R20. La ausencia de CCM2 compromete la activación de ERK1/2.

Figura R21. CCM2 no interacciona de manera directa con componentes de la vía ERK1/2.

Figura R22. La pérdida de función de CCM2 en ESCs altera los niveles transcripcionales de reguladores positivos y negativos que participan en el control de la activación de ERK1/2.

Figura R23. Análisis de similitud entre *Ccm2* KO y la inhibición de MEK o GSK3β en ESCs.

Figura R24. La ausencia de CCM2 afecta a la expresión de dianas transcripcionales de ERK2.

Figura R25. La inhibición de ERK1/2 contribuye a gran parte de la desregulación transcripcional mediada por la ausencia de CCM2.

Figura R26. La ausencia de CCM2 altera la expresión de genes diana de PRC2.

Figura R27. CCM2 participa en la regulación transcripcional de genes bivalentes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figura R28. La ausencia de actividad catalítica de PRC2 explica parte de la desregulación transcripcional de las ESCs KO para *Ccm2*.

Figura R29. La ausencia de CCM2 altera la abundancia de 5-mC y 5-hmC en ADN.

Figura R30. La eliminación de CCM2 modifica los niveles globales de DNMTs y TETs.

Figura R31. DNMT3B, TET1 y TET2 se unen a genes desregulados transcripcionalmente en ausencia de CCM2.

Figura R32. DNMT3B, TET1 y TET2 se encuentran presentes en una fracción de genes desregulados tras la pérdida de CCM2.

Figura R33. DNMT3B, TET1 y TET2 participan en el control de la regulación transcripcional de genes desregulados en ausencia de CCM2.

Figura R34. Identificación del interactoma de CCM2 en ESCs.

Figura R35. Análisis del efecto de la pérdida de CCM2 sobre la expresión de los genes que codifican a sus interactores.

Figura R36. Análisis de los procesos y fenotipos en los que intervienen los interactores de CCM2.

Figura R37. Cribado *in silico* de CCM2 e interactores en la regulación del gen *naïve Rex1*.

Figura R38. CCM2 interacciona con varios complejos que regulan la cromatina.

Figura R39. La ausencia de CCM2 afecta a regiones potenciadoras y promotoras en genes desregulados.

Figura R40. La pérdida de función CCM altera programas transcripcionales de diferenciación celular en ESCs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figura R41. Las proteínas CCM son necesarias durante la formación de EBs.

Figura R42. Las CCM son necesarias para una correcta formación de gastruloides.

Figura R43. Determinación del protocolo óptimo de diferenciación de ESCs hacia EPCs.

Figura R44. Establecimiento de una firma de expresión génica para EPCs y validación de las EPCs generadas a partir de ESCs.

Figura R45. La ausencia de las CCM impide una correcta formación de EPCs.

Figura R46. Identificación de genes comunes desregulados en ausencia de cada proteína CCM relacionados con la función endotelial.

Figura R47. Identificación de fármacos que reviertan los defectos de las EPCs.

Tabla R1. La ausencia de las CCM altera la expresión de genes diana de factores de transcripción relacionados con la función endotelial.

Figura D1. Representación esquemática simplificada de la relevancia de las proteínas CCM en pluripotencia y especificación endotelial.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

11. BIBLIOGRAFÍA

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.Condic ML. Totipotency: what it is and what it is not. Stem cells and development. 2014;23(8):796-812.

2.Wu D, Dean J. Maternal factors regulating preimplantation development in mice. Current topics in developmental biology. 2020;140:317-40.

3.Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nature reviews Genetics. 2019;20(4):221-34.

4.Cockburn K, Rossant J. Making the blastocyst: lessons from the mouse. The Journal of clinical investigation. 2010;120(4):995-1003.

5.White MD, Plachta N. How adhesion forms the early mammalian embryo. Current topics in developmental biology. 2015;112:1-17.

6.Johnson MH, McConnell JM. Lineage allocation and cell polarity during mouse embryogenesis. Seminars in cell & developmental biology. 2004;15(5):583-97.

7.Nishioka N, Inoue K, Adachi K, Kiyonari H, Ota M, Ralston A, et al. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Developmental cell. 2009;16(3):398-410.

8.Yagi R, Kohn MJ, Karavanova I, Kaneko KJ, Vullhorst D, DePamphilis ML, et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development (Cambridge, England). 2007;134(21):3827-36.

9.Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development (Cambridge, England). 2005;132(9):2093-102.

10.Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development (Cambridge, England). 2010;137(3):395-403.

11.Xenopoulos P, Kang M, Hadjantonakis AK. Cell lineage allocation within the inner cell mass of the mouse blastocyst. Results and problems in cell differentiation. 2012;55:185-202.

12.Karasek C, Ashry M, Driscoll CS, Knott JG. A tale of two cell-fates: role of the Hippo signaling pathway and transcription factors in early lineage formation in mouse preimplantation embryos. Molecular human reproduction. 2020;26(9):653-64.

13.Geng T, Zhang D, Jiang W. Epigenetic Regulation of Transition Among Different Pluripotent States: Concise Review. Stem cells (Dayton, Ohio). 2019;37(11):1372-80.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

14.Yamanaka Y, Lanner F, Rossant J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development (Cambridge, England). 2010;137(5):715-24.

15.Kim SM, Kim JS. A Review of Mechanisms of Implantation. Development & reproduction. 2017;21(4):351-9.

16.Bénazéraf B, Pourquié O. Formation and segmentation of the vertebrate body axis. Annual review of cell and developmental biology. 2013;29:1-26.

17.Nichols J, Smith A. Naive and primed pluripotent states. Cell stem cell. 2009;4(6):487-92.

18.Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature. 2007;448(7150):191-5.

19.Kinoshita M, Barber M, Mansfield W, Cui Y, Spindlow D, Stirparo GG, et al. Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. Cell stem cell. 2021;28(3):453-71.e8.

20.Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature. 2007;448(7150):196-9.

21.Rossant J. Stem cells and early lineage development. Cell. 2008;132(4):527-31.

22.Nichols J, Smith A. Pluripotency in the embryo and in culture. Cold Spring Harbor perspectives in biology. 2012;4(8):a008128.

23.Boutourlinsky K, Allègre N, Chazaud C. Ex Vivo Culture for Preimplantation Mouse Embryo to Analyze Pluripotency. Methods in molecular biology (Clifton, NJ). 2021;2214:1-10.

24.Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America. 1981;78(12):7634-8.

25.Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981;292(5819):154-6.

26.Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nature reviews Molecular cell biology. 2016;17(3):155-69.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

27.Devika AS, Wruck W, Adjaye J, Sudheer S. The quest for pluripotency: a comparative analysis across mammalian species. Reproduction (Cambridge, England). 2019;158(3):R97-r111.

28.Huang Y, Osorno R, Tsakiridis A, Wilson V. In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell reports. 2012;2(6):1571-8.

29.Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008;453(7194):519-23.

30.Nusse R, Clevers H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. Cell. 2017;169(6):985-99.

31.MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. Developmental cell. 2009;17(1):9-26.

32.Chang F, Steelman LS, Lee JT, Shelton JG, Navolanic PM, Blalock WL, et al. Signal transduction mediated by the Ras/Raf/MEK/ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention. Leukemia. 2003;17(7):1263-93.

33.Hayashi K, de Sousa Lopes SMC, Tang F, Lao K, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell stem cell. 2008;3(4):391-401.

34.Guo G, Pinello L, Han X, Lai S, Shen L, Lin TW, et al. Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis. Cell reports. 2016;14(4):956-65.

35.Neagu A, van Genderen E, Escudero I, Verwegen L, Kurek D, Lehmann J, et al. In vitro capture and characterization of embryonic rosette-stage pluripotency between naive and primed states. Nature cell biology. 2020;22(5):534-45.

36.Shahbazi MN, Scialdone A, Skorupska N, Weberling A, Recher G, Zhu M, et al. Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. Nature. 2017;552(7684):239-43.

37.Smith A. Formative pluripotency: the executive phase in a developmental continuum. Development (Cambridge, England). 2017;144(3):365-73.

38.Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development (Cambridge, England). 2009;136(7):1063-9.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

39.Han DW, Tapia N, Joo JY, Greber B, Araúzo-Bravo MJ, Bernemann C, et al. Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. Cell. 2010;143(4):617-27.

40.Wang X, Xiang Y, Yu Y, Wang R, Zhang Y, Xu Q, et al. Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. Cell research. 2021;31(5):526-41.

41.Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. The Journal of biological chemistry. 2005;280(7):5307-17.

42.Orkin SH, Wang J, Kim J, Chu J, Rao S, Theunissen TW, et al. The transcriptional network controlling pluripotency in ES cells. Cold Spring Harbor symposia on quantitative biology. 2008;73:195-202.

43.Ng HH, Surani MA. The transcriptional and signalling networks of pluripotency. Nature cell biology. 2011;13(5):490-6.

44.Niwa H. The pluripotency transcription factor network at work in reprogramming. Current opinion in genetics & development. 2014;28:25-31.

45.Wray J, Kalkan T, Smith AG. The ground state of pluripotency. Biochemical Society transactions. 2010;38(4):1027-32.

46.Azami T, Matsumoto K, Jeon H, Waku T, Muratani M, Niwa H, et al. Klf5 suppresses ERK signaling in mouse pluripotent stem cells. PloS one. 2018;13(11):e0207321.

47.Yeo JC, Ng HH. The transcriptional regulation of pluripotency. Cell research. 2013;23(1):20-32.

48.Choi HW, Joo JY, Hong YJ, Kim JS, Song H, Lee JW, et al. Distinct Enhancer Activity of Oct4 in Naive and Primed Mouse Pluripotency. Stem cell reports. 2016;7(5):911-26.

49.Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell stem cell. 2014;14(6):838-53.

50.Yang SH, Kalkan T, Morissroe C, Marks H, Stunnenberg H, Smith A, et al. Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell reports. 2014;7(6):1968-81.

51.Messmer T, von Meyenn F, Savino A, Santos F, Mohammed H, Lun ATL, et al. Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. Cell reports. 2019;26(4):815-24.e4.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

52.Yagi M, Yamanaka S, Yamada Y. Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. Laboratory investigation; a journal of technical methods and pathology. 2017;97(10):1133-41.

53.Yu L, Wei Y, Sun HX, Mahdi AK, Pinzon Arteaga CA, Sakurai M, et al. Derivation of Intermediate Pluripotent Stem Cells Amenable to Primordial Germ Cell Specification. Cell stem cell. 2021;28(3):550-67.e12.

54.Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, et al. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell stem cell. 2014;14(6):854-63.

55.Hayashi K, Surani MA. Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. Development (Cambridge, England). 2009;136(21):3549-56.

56.Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nature reviews Molecular cell biology. 2019;20(10):590-607.

57.Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247-57.

58.Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proceedings of the National Academy of Sciences of the United States of America. 2011;108(22):9055-9.

59.He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science (New York, NY). 2011;333(6047):1303-7.

60.Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714-7.

61.Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science (New York, NY). 2007;317(5845):1760-4.

62.Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129-33.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

63.Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (New York, NY). 2011;333(6047):1300-3.

64.An J, Rao A, Ko M. TET family dioxygenases and DNA demethylation in stem cells and cancers. Experimental & molecular medicine. 2017;49(4):e323.

65.Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic acids research. 2012;40(11):4841-9.

66.Otani J, Kimura H, Sharif J, Endo TA, Mishima Y, Kawakami T, et al. Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. PloS one. 2013;8(12):e82961.

67.Weber AR, Krawczyk C, Robertson AB, Kuśnierczyk A, Vågbø CB, Schuermann D, et al. Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nature communications. 2016;7:10806.

68.Garcia-Outeiral V, de la Parte C, Fidalgo M, Guallar D. The Complexity of TET2 Functions in Pluripotency and Development. Frontiers in cell and developmental biology. 2020;8:630754.

69.Sim YJ, Kim MS, Nayfeh A, Yun YJ, Kim SJ, Park KT, et al. 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. Stem cell reports. 2017;8(5):1312-28.

70.Shanak S, Helms V. DNA methylation and the core pluripotency network. Developmental biology. 2020;464(2):145-60.

71.Betto RM, Diamante L, Perrera V, Audano M, Rapelli S, Lauria A, et al. Metabolic control of DNA methylation in naive pluripotent cells. Nature genetics. 2021;53(2):215-29.

72.Cao K, Collings CK, Morgan MA, Marshall SA, Rendleman EJ, Ozark PA, et al. An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. Science advances. 2018;4(1):eaap8747.

73.Fidalgo M, Huang X, Guallar D, Sanchez-Priego C, Valdes VJ, Saunders A, et al. Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent States. Cell stem cell. 2016;19(3):355-69.

74.Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, et al. Naive pluripotency is associated with global DNA hypomethylation. Nature structural & molecular biology. 2013;20(3):311-6.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

75.Kalkan T, Olova N, Roode M, Mulas C, Lee HJ, Nett I, et al. Tracking the embryonic stem cell transition from ground state pluripotency. Development (Cambridge, England). 2017;144(7):1221-34.

76.Mulholland CB, Traube FR, Ugur E, Parsa E, Eckl EM, Schönung M, et al. Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. Scientific reports. 2020;10(1):12066.

77.Parry A, Rulands S, Reik W. Active turnover of DNA methylation during cell fate decisions. Nature reviews Genetics. 2021;22(1):59-66.

78.Gökbuget D, Blelloch R. Epigenetic control of transcriptional regulation in pluripotency and early differentiation. Development (Cambridge, England). 2019;146(19).

79.Jambhekar A, Dhall A, Shi Y. Roles and regulation of histone methylation in animal development. Nature reviews Molecular cell biology. 2019;20(10):625-41.

80.Di Croce L, Helin K. Transcriptional regulation by Polycomb group proteins. Nature structural & molecular biology. 2013;20(10):1147-55.

81.Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 2006;441(7091):349-53.

82.O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. The polycomb-group gene Ezh2 is required for early mouse development. Molecular and cellular biology. 2001;21(13):4330-6.

83.Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, et al. Role of histone H2A ubiquitination in Polycomb silencing. Nature. 2004;431(7010):873-8.

84.Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (New York, NY). 2002;298(5595):1039-43.

85.Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Molecular cell. 2012;45(3):344-56.

86.Zhu Y, Dong L, Wang C, Hao K, Wang J, Zhao L, et al. Functional redundancy among Polycomb complexes in maintaining the pluripotent state of embryonic stem cells. Stem cell reports. 2022;17(5):1198-214.

87.van Mierlo G, Veenstra GJC, Vermeulen M, Marks H. The Complexity of PRC2 Subcomplexes. Trends in cell biology. 2019;29(8):660-71.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

88.Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LLP, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell. 2014;157(6):1445-59.

89.Morey L, Aloia L, Cozzuto L, Benitah SA, Di Croce L. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell reports. 2013;3(1):60-9.

90.Creppe C, Palau A, Malinverni R, Valero V, Buschbeck M. A Cbx8-containing polycomb complex facilitates the transition to gene activation during ES cell differentiation. PLoS genetics. 2014;10(12):e1004851.

91.Schaaf CA, Misulovin Z, Gause M, Koenig A, Gohara DW, Watson A, et al. Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS genetics. 2013;9(6):e1003560.

92.Pherson M, Misulovin Z, Gause M, Mihindukulasuriya K, Swain A, Dorsett D. Polycomb repressive complex 1 modifies transcription of active genes. Science advances. 2017;3(8):e1700944.

93.Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. Science (New York, NY). 2016;352(6290):aad9780.

94.Bledau AS, Schmidt K, Neumann K, Hill U, Ciotta G, Gupta A, et al. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. Development (Cambridge, England). 2014;141(5):1022-35.

95.Glaser S, Lubitz S, Loveland KL, Ohbo K, Robb L, Schwenk F, et al. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics & chromatin. 2009;2(1):5.

96.Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell. 2005;121(6):859-72.

97.Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nature structural & molecular biology. 2006;13(9):852-4.

98.Xiang Y, Zhang Y, Xu Q, Zhou C, Liu B, Du Z, et al. Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. Nature genetics. 2020;52(1):95-105.

99.Shema E, Jones D, Shoresh N, Donohue L, Ram O, Bernstein BE. Single-molecule decoding of combinatorially modified nucleosomes. Science (New York, NY). 2016;352(6286):717-21.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

100.Voigt P, LeRoy G, Drury WJ, 3rd, Zee BM, Son J, Beck DB, et al. Asymmetrically modified nucleosomes. Cell. 2012;151(1):181-93.

101.Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. Genes & development. 2013;27(12):1318-38.

102.Harikumar A, Meshorer E. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. EMBO reports. 2015;16(12):1609-19.

103.van Mierlo G, Dirks RAM, De Clerck L, Brinkman AB, Huth M, Kloet SL, et al. Integrative Proteomic Profiling Reveals PRC2-Dependent Epigenetic Crosstalk Maintains Ground-State Pluripotency. Cell stem cell. 2019;24(1):123-37.e8.

104.Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, et al. The transcriptional and epigenomic foundations of ground state pluripotency. Cell. 2012;149(3):590-604.

105.Tee WW, Shen SS, Oksuz O, Narendra V, Reinberg D. Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. Cell. 2014;156(4):678-90.

106.Chamberlain SJ, Yee D, Magnuson T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem cells (Dayton, Ohio). 2008;26(6):1496-505.

107.Endoh M, Endo TA, Endoh T, Fujimura Y, Ohara O, Toyoda T, et al. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development (Cambridge, England). 2008;135(8):1513-24.

108.Wang C, Lee JE, Lai B, Macfarlan TS, Xu S, Zhuang L, et al. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(42):11871-6.

109.Lubitz S, Glaser S, Schaft J, Stewart AF, Anastassiadis K. Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. Molecular biology of the cell. 2007;18(6):2356-66.

110.Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, Singh S, et al. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development (Cambridge, England). 2014;141(3):526-37.

111.Mas G, Blanco E, Ballaré C, Sansó M, Spill YG, Hu D, et al. Promoter bivalency favors an open chromatin architecture in embryonic stem cells. Nature genetics. 2018;50(10):1452-62.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

112.Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Di Croce L. The Bivalent Genome: Characterization, Structure, and Regulation. Trends in genetics : TIG. 2020;36(2):118-31.

113.Hirai H, Karian P, Kikyo N. Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. The Biochemical journal. 2011;438(1):11-23.

114.Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW. Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene. 2002;285(1-2):1-24.

115.Do DV, Ueda J, Messerschmidt DM, Lorthongpanich C, Zhou Y, Feng B, et al. A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes & development. 2013;27(12):1378-90.

116.Ho L, Miller EL, Ronan JL, Ho WQ, Jothi R, Crabtree GR. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. Nature cell biology. 2011;13(8):903-13.

117.Niwa H, Ogawa K, Shimosato D, Adachi K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature. 2009;460(7251):118-22.

118.Ye S, Li P, Tong C, Ying QL. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. The EMBO journal. 2013;32(19):2548-60.

119.Paling NR, Wheadon H, Bone HK, Welham MJ. Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. The Journal of biological chemistry. 2004;279(46):48063-70.

120.Yu JS, Cui W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination. Development (Cambridge, England). 2016;143(17):3050-60.

121.Bechard M, Dalton S. Subcellular localization of glycogen synthase kinase 3beta controls embryonic stem cell self-renewal. Molecular and cellular biology. 2009;29(8):2092-104.

122.Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, Dalton S. LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development (Cambridge, England). 2005;132(5):885-96.

123.Wu Y, Liu F, Liu Y, Liu X, Ai Z, Guo Z, et al. GSK3 inhibitors CHIR99021 and 6-bromoindirubin-3'-oxime inhibit microRNA maturation in mouse embryonic stem cells. Scientific reports. 2015;5:8666.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

124.Mossahebi-Mohammadi M, Quan M, Zhang JS, Li X. FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency. Frontiers in cell and developmental biology. 2020;8:79.

125.Ying QL, Nichols J, Chambers I, Smith A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell. 2003;115(3):281-92.

126.Morikawa M, Koinuma D, Mizutani A, Kawasaki N, Holmborn K, Sundqvist A, et al. BMP Sustains Embryonic Stem Cell Self-Renewal through Distinct Functions of Different Krüppel-like Factors. Stem cell reports. 2016;6(1):64-73.

127.de Caestecker M. The transforming growth factor-beta superfamily of receptors. Cytokine & growth factor reviews. 2004;15(1):1-11.

128.Horbelt D, Denkis A, Knaus P. A portrait of Transforming Growth Factor ß superfamily signalling: Background matters. The international journal of biochemistry & cell biology. 2012;44(3):469-74.

129.Ruzinova MB, Benezra R. Id proteins in development, cell cycle and cancer. Trends in cell biology. 2003;13(8):410-8.

130.Li Z, Fei T, Zhang J, Zhu G, Wang L, Lu D, et al. BMP4 Signaling Acts via dual-specificity phosphatase 9 to control ERK activity in mouse embryonic stem cells. Cell stem cell. 2012;10(2):171-82.

131.Chng Z, Vallier L, Pedersen R. Activin/nodal signaling and pluripotency. Vitamins and hormones. 2011;85:39-58.

132.Wiese KE, Nusse R, van Amerongen R. Wnt signalling: conquering complexity. Development (Cambridge, England). 2018;145(12).

133.Li J, Li J, Chen B. Oct4 was a novel target of Wnt signaling pathway. Molecular and cellular biochemistry. 2012;362(1-2):233-40.

134.Kelly KF, Ng DY, Jayakumaran G, Wood GA, Koide H, Doble BW. ß-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. Cell stem cell. 2011;8(2):214-27.

135.Takao Y, Yokota T, Koide H. Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. Biochemical and biophysical research communications. 2007;353(3):699-705.

136.Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes & development. 2008;22(6):746-55.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

137.Yi F, Pereira L, Merrill BJ. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. Stem cells (Dayton, Ohio). 2008;26(8):1951-60.

138.Martello G, Sugimoto T, Diamanti E, Joshi A, Hannah R, Ohtsuka S, et al. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell stem cell. 2012;11(4):491-504.

139.Geary L, LaBonne C. FGF mediated MAPK and PI3K/Akt Signals make distinct contributions to pluripotency and the establishment of Neural Crest. eLife. 2018;7.

140.Ma X, Chen H, Chen L. A dual role of Erk signaling in embryonic stem cells. Experimental hematology. 2016;44(3):151-6.

141.Lanner F, Rossant J. The role of FGF/Erk signaling in pluripotent cells. Development (Cambridge, England). 2010;137(20):3351-60.

142.Nett IR, Mulas C, Gatto L, Lilley KS, Smith A. Negative feedback via RSK modulates Erk-dependent progression from naïve pluripotency. EMBO reports. 2018;19(8).

143.Sarabipour S, Hristova K. Mechanism of FGF receptor dimerization and activation. Nature communications. 2016;7:10262.

144.Margarit SM, Sondermann H, Hall BE, Nagar B, Hoelz A, Pirruccello M, et al. Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. Cell. 2003;112(5):685-95.

145.Shoni M, Lui KO, Vavvas DG, Muto MG, Berkowitz RS, Vlahos N, et al. Protein kinases and associated pathways in pluripotent state and lineage differentiation. Current stem cell research & therapy. 2014;9(5):366-87.

146.Hamilton WB, Mosesson Y, Monteiro RS, Emdal KB, Knudsen TE, Francavilla C, et al. Dynamic lineage priming is driven via direct enhancer regulation by ERK. Nature. 2019;575(7782):355-60.

147.Brumbaugh J, Russell JD, Yu P, Westphall MS, Coon JJ, Thomson JA. NANOG is multiply phosphorylated and directly modified by ERK2 and CDK1 in vitro. Stem cell reports. 2014;2(1):18-25.

148.Dhaliwal NK, Miri K, Davidson S, Tamim El Jarkass H, Mitchell JA. KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency. Stem cell reports. 2018;10(4):1308-23.

149.Kim MO, Kim SH, Cho YY, Nadas J, Jeong CH, Yao K, et al. ERK1 and ERK2 regulate embryonic stem cell self-renewal through phosphorylation of Klf4. Nature structural & molecular biology. 2012;19(3):283-90.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

150.Spelat R, Ferro F, Curcio F. Serine 111 phosphorylation regulates OCT4A protein subcellular distribution and degradation. The Journal of biological chemistry. 2012;287(45):38279-88.

151.Ma C, Karwacki-Neisius V, Tang H, Li W, Shi Z, Hu H, et al. Nono, a Bivalent Domain Factor, Regulates Erk Signaling and Mouse Embryonic Stem Cell Pluripotency. Cell reports. 2016;17(4):997-1007.

152.Sun HL, Zhu AC, Gao Y, Terajima H, Fei Q, Liu S, et al. Stabilization of ERK-Phosphorylated METTL3 by USP5 Increases m(6)A Methylation. Molecular cell. 2020;80(4):633-47.e7.

153.Kubo A, Shinozaki K, Shannon JM, Kouskoff V, Kennedy M, Woo S, et al. Development of definitive endoderm from embryonic stem cells in culture. Development (Cambridge, England). 2004;131(7):1651-62.

154.Yasunaga M, Tada S, Torikai-Nishikawa S, Nakano Y, Okada M, Jakt LM, et al. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. Nature biotechnology. 2005;23(12):1542-50.

155.Takenaga M, Fukumoto M, Hori Y. Regulated Nodal signaling promotes differentiation of the definitive endoderm and mesoderm from ES cells. Journal of cell science. 2007;120(Pt 12):2078-90.

156.Sakurai H, Inami Y, Tamamura Y, Yoshikai T, Sehara-Fujisawa A, Isobe K. Bidirectional induction toward paraxial mesodermal derivatives from mouse ES cells in chemically defined medium. Stem cell research. 2009;3(2-3):157-69.

157.Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, van Oudenaarden A. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nature communications. 2017;8(1):1096.

158.Rochette-Egly C. Retinoic acid signaling and mouse embryonic stem cell differentiation: Cross talk between genomic and non-genomic effects of RA. Biochimica et biophysica acta. 2015;1851(1):66-75.

159.Era T, Izumi N, Hayashi M, Tada S, Nishikawa S, Nishikawa S. Multiple mesoderm subsets give rise to endothelial cells, whereas hematopoietic cells are differentiated only from a restricted subset in embryonic stem cell differentiation culture. Stem cells (Dayton, Ohio). 2008;26(2):401-11.

160.Shalaby F, Rossant J, Yamaguchi TP, Gertsenstein M, Wu XF, Breitman ML, et al. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. Nature. 1995;376(6535):62-6.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

161.Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, et al. Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. Nature. 1996;380(6573):435-9.

162.Augustin HG, Koh GY. Organotypic vasculature: From descriptive heterogeneity to functional pathophysiology. Science (New York, NY). 2017;357(6353).

163.Niklason L, Dai G. Arterial Venous Differentiation for Vascular Bioengineering. Annual review of biomedical engineering. 2018;20:431-47.

164.Chan AC, Li DY, Berg MJ, Whitehead KJ. Recent insights into cerebral cavernous malformations: animal models of CCM and the human phenotype. The FEBS journal. 2010;277(5):1076-83.

165.Guzeloglu-Kayisli O, Amankulor NM, Voorhees J, Luleci G, Lifton RP, Gunel M. KRIT1/cerebral cavernous malformation 1 protein localizes to vascular endothelium, astrocytes, and pyramidal cells of the adult human cerebral cortex. Neurosurgery. 2004;54(4):943-9; discussion 9.

166.Glading A, Han J, Stockton RA, Ginsberg MH. KRIT-1/CCM1 is a Rap1 effector that regulates endothelial cell cell junctions. The Journal of cell biology. 2007;179(2):247-54.

167.Draheim KM, Fisher OS, Boggon TJ, Calderwood DA. Cerebral cavernous malformation proteins at a glance. Journal of cell science. 2014;127(Pt 4):701-7.

168.Fisher OS, Boggon TJ. Signaling pathways and the cerebral cavernous malformations proteins: lessons from structural biology. Cellular and molecular life sciences : CMLS. 2014;71(10):1881-92.

169.Riolo G, Ricci C, Battistini S. Molecular Genetic Features of Cerebral Cavernous Malformations (CCM) Patients: An Overall View from Genes to Endothelial Cells. Cells. 2021;10(3).

170.Zawistowski JS, Serebriiskii IG, Lee MF, Golemis EA, Marchuk DA. KRIT1 association with the integrin-binding protein ICAP-1: a new direction in the elucidation of cerebral cavernous malformations (CCM1) pathogenesis. Human molecular genetics. 2002;11(4):389-96.

171.Zawistowski JS, Stalheim L, Uhlik MT, Abell AN, Ancrile BB, Johnson GL, et al. CCM1 and CCM2 protein interactions in cell signaling: implications for cerebral cavernous malformations pathogenesis. Human molecular genetics. 2005;14(17):2521-31.

172.Stiegler AL, Zhang R, Liu W, Boggon TJ. Structural determinants for binding of sorting nexin 17 (SNX17) to the cytoplasmic adaptor protein Krev interaction trapped 1 (KRIT1). The Journal of biological chemistry. 2014;289(36):25362-73.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

173.Li X, Zhang R, Draheim KM, Liu W, Calderwood DA, Boggon TJ. Structural basis for small G protein effector interaction of Ras-related protein 1 (Rap1) and adaptor protein Krev interaction trapped 1 (KRIT1). The Journal of biological chemistry. 2012;287(26):22317-27.

174.Gingras AR, Puzon-McLaughlin W, Ginsberg MH. The structure of the ternary complex of Krev interaction trapped 1 (KRIT1) bound to both the Rap1 GTPase and the heart of glass (HEG1) cytoplasmic tail. The Journal of biological chemistry. 2013;288(33):23639-49.

175.Seker A, Pricola KL, Guclu B, Ozturk AK, Louvi A, Gunel M. CCM2 expression parallels that of CCM1. Stroke. 2006;37(2):518-23.

176.Fisher OS, Zhang R, Li X, Murphy JW, Demeler B, Boggon TJ. Structural studies of cerebral cavernous malformations 2 (CCM2) reveal a folded helical domain at its C-terminus. FEBS letters. 2013;587(3):272-7.

177.Faurobert E, Albiges-Rizo C. Recent insights into cerebral cavernous malformations: a complex jigsaw puzzle under construction. The FEBS journal. 2010;277(5):1084-96.

178.Whitehead KJ, Chan AC, Navankasattusas S, Koh W, London NR, Ling J, et al. The cerebral cavernous malformation signaling pathway promotes vascular integrity via Rho GTPases. Nature medicine. 2009;15(2):177-84.

179.Fisher OS, Deng H, Liu D, Zhang Y, Wei R, Deng Y, et al. Structure and vascular function of MEKK3-cerebral cavernous malformations 2 complex. Nature communications. 2015;6:7937.

180.Tanriover G, Boylan AJ, Diluna ML, Pricola KL, Louvi A, Gunel M. PDCD10, the gene mutated in cerebral cavernous malformation 3, is expressed in the neurovascular unit. Neurosurgery. 2008;62(4):930-8; discussion 8.

181.Fidalgo M, Fraile M, Pires A, Force T, Pombo C, Zalvide J. CCM3/PDCD10 stabilizes GCKIII proteins to promote Golgi assembly and cell orientation. Journal of cell science. 2010;123(Pt 8):1274-84.

182.Ding J, Wang X, Li DF, Hu Y, Zhang Y, Wang DC. Crystal structure of human programmed cell death 10 complexed with inositol-(1,3,4,5)-tetrakisphosphate: a novel adaptor protein involved in human cerebral cavernous malformation. Biochemical and biophysical research communications. 2010;399(4):587-92.

183.Goudreault M, D'Ambrosio LM, Kean MJ, Mullin MJ, Larsen BG, Sanchez A, et al. A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. Molecular & cellular proteomics : MCP. 2009;8(1):157-71.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

184.Ceccarelli DF, Laister RC, Mulligan VK, Kean MJ, Goudreault M, Scott IC, et al. CCM3/PDCD10 heterodimerizes with germinal center kinase III (GCKIII) proteins using a mechanism analogous to CCM3 homodimerization. The Journal of biological chemistry. 2011;286(28):25056-64.

185.Padarti A, Zhang J. Recent advances in cerebral cavernous malformation research. Vessel plus. 2018;2.

186.Draheim KM, Li X, Zhang R, Fisher OS, Villari G, Boggon TJ, et al. CCM2-CCM3 interaction stabilizes their protein expression and permits endothelial network formation. The Journal of cell biology. 2015;208(7):987-1001.

187.Dibble CF, Horst JA, Malone MH, Park K, Temple B, Cheeseman H, et al. Defining the functional domain of programmed cell death 10 through its interactions with phosphatidylinositol-3,4,5-trisphosphate. PloS one. 2010;5(7):e11740.

188.He Y, Zhang H, Yu L, Gunel M, Boggon TJ, Chen H, et al. Stabilization of VEGFR2 signaling by cerebral cavernous malformation 3 is critical for vascular development. Science signaling. 2010;3(116):ra26.

189.Mardakheh FK, Self A, Marshall CJ. RHO binding to FAM65A regulates Golgi reorientation during cell migration. Journal of cell science. 2016;129(24):4466-79.

190.Voss K, Stahl S, Schleider E, Ullrich S, Nickel J, Mueller TD, et al. CCM3 interacts with CCM2 indicating common pathogenesis for cerebral cavernous malformations. Neurogenetics. 2007;8(4):249-56.

191.Hilder TL, Malone MH, Bencharit S, Colicelli J, Haystead TA, Johnson GL, et al. Proteomic identification of the cerebral cavernous malformation signaling complex. Journal of proteome research. 2007;6(11):4343-55.

192.Stockton RA, Shenkar R, Awad IA, Ginsberg MH. Cerebral cavernous malformations proteins inhibit Rho kinase to stabilize vascular integrity. The Journal of experimental medicine. 2010;207(4):881-96.

193.Crose LES, Hilder TL, Sciaky N, Johnson GL. Cerebral cavernous malformation 2 protein promotes smad ubiquitin regulatory factor 1-mediated RhoA degradation in endothelial cells. The Journal of biological chemistry. 2009;284(20):13301-5.

194.Ma X, Zhao H, Shan J, Long F, Chen Y, Chen Y, et al. PDCD10 interacts with Ste20-related kinase MST4 to promote cell growth and transformation via modulation of the ERK pathway. Molecular biology of the cell. 2007;18(6):1965-78.

195.Fidalgo M, Guerrero A, Fraile M, Iglesias C, Pombo CM, Zalvide J. Adaptor protein cerebral cavernous malformation 3 (CCM3) mediates phosphorylation of the cytoskeletal proteins ezrin/radixin/moesin by mammalian Ste20-4 to protect cells from oxidative stress. The Journal of biological chemistry. 2012;287(14):11556-65.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

196.Jahn R, Südhof TC. Membrane fusion and exocytosis. Annual review of biochemistry. 1999;68:863-911.

197.Zhao H, Mleynek TM, Li DY. Dysregulated exocytosis of angiopoietin-2 drives cerebral cavernous malformation. Nature medicine. 2016;22(9):971-3.

198.Glading AJ, Ginsberg MH. Rap1 and its effector KRIT1/CCM1 regulate beta-catenin signaling. Disease models & mechanisms. 2010;3(1-2):73-83.

199.de Kreuk BJ, Gingras AR, Knight JD, Liu JJ, Gingras AC, Ginsberg MH. Heart of glass anchors Rasip1 at endothelial cell-cell junctions to support vascular integrity. eLife. 2016;5:e11394.

200.Bauer HC, Krizbai IA, Bauer H, Traweger A. "You Shall Not Pass"-tight junctions of the blood brain barrier. Frontiers in neuroscience. 2014;8:392.

201.Ivanov DB, Philippova MP, Tkachuk VA. Structure and functions of classical cadherins. Biochemistry Biokhimiia. 2001;66(10):1174-86.

202.Dejana E. Endothelial cell-cell junctions: happy together. Nature reviews Molecular cell biology. 2004;5(4):261-70.

203.Boulday G, Rudini N, Maddaluno L, Blécon A, Arnould M, Gaudric A, et al. Developmental timing of CCM2 loss influences cerebral cavernous malformations in mice. The Journal of experimental medicine. 2011;208(9):1835-47.

204.Brütsch R, Liebler SS, Wüstehube J, Bartol A, Herberich SE, Adam MG, et al. Integrin cytoplasmic domain-associated protein-1 attenuates sprouting angiogenesis. Circulation research. 2010;107(5):592-601.

205.Wüstehube J, Bartol A, Liebler SS, Brütsch R, Zhu Y, Felbor U, et al. Cerebral cavernous malformation protein CCM1 inhibits sprouting angiogenesis by activating DELTA-NOTCH signaling. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(28):12640-5.

206.Millon-Frémillon A, Bouvard D, Grichine A, Manet-Dupé S, Block MR, Albiges-Rizo C. Cell adaptive response to extracellular matrix density is controlled by ICAP-1-dependent beta1-integrin affinity. The Journal of cell biology. 2008;180(2):427-41.

207.Bouvard D, Vignoud L, Dupé-Manet S, Abed N, Fournier HN, Vincent-Monegat C, et al. Disruption of focal adhesions by integrin cytoplasmic domain-associated protein-1 alpha. The Journal of biological chemistry. 2003;278(8):6567-74.

208.Singh P, Carraher C, Schwarzbauer JE. Assembly of fibronectin extracellular matrix. Annual review of cell and developmental biology. 2010;26:397-419.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

209.Faurobert E, Rome C, Lisowska J, Manet-Dupé S, Boulday G, Malbouyres M, et al. CCM1-ICAP-1 complex controls β1 integrin-dependent endothelial contractility and fibronectin remodeling. The Journal of cell biology. 2013;202(3):545-61.

210.Zhou Z, Rawnsley DR, Goddard LM, Pan W, Cao XJ, Jakus Z, et al. The cerebral cavernous malformation pathway controls cardiac development via regulation of endocardial MEKK3 signaling and KLF expression. Developmental cell. 2015;32(2):168-80.

211.Wang X, Hou Y, Deng K, Zhang Y, Wang DC, Ding J. Structural Insights into the Molecular Recognition between Cerebral Cavernous Malformation 2 and Mitogen-Activated Protein Kinase Kinase Kinase 3. Structure (London, England : 1993). 2015;23(6):1087-96.

212.Yang J, Boerm M, McCarty M, Bucana C, Fidler IJ, Zhuang Y, et al. Mekk3 is essential for early embryonic cardiovascular development. Nature genetics. 2000;24(3):309-13.

213.Cullere X, Plovie E, Bennett PM, MacRae CA, Mayadas TN. The cerebral cavernous malformation proteins CCM2L and CCM2 prevent the activation of the MAP kinase MEKK3. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(46):14284-9.

214.Zhou Z, Tang AT, Wong WY, Bamezai S, Goddard LM, Shenkar R, et al. Cerebral cavernous malformations arise from endothelial gain of MEKK3-KLF2/4 signalling. Nature. 2016;532(7597):122-6.

215.Petit N, Blécon A, Denier C, Tournier-Lasserve E. Patterns of expression of the three cerebral cavernous malformation (CCM) genes during embryonic and postnatal brain development. Gene expression patterns : GEP. 2006;6(5):495-503.

216.Wang Y, Liu H, Zhang Y, Ma D. cDNA cloning and expression of an apoptosis-related gene, humanTFAR15 gene. Science in China Series C, Life sciences. 1999;42(3):323-9.

217.Whitehead KJ, Plummer NW, Adams JA, Marchuk DA, Li DY. Ccm1 is required for arterial morphogenesis: implications for the etiology of human cavernous malformations. Development (Cambridge, England). 2004;131(6):1437-48.

218.Boulday G, Blécon A, Petit N, Chareyre F, Garcia LA, Niwa-Kawakita M, et al. Tissue-specific conditional CCM2 knockout mice establish the essential role of endothelial CCM2 in angiogenesis: implications for human cerebral cavernous malformations. Disease models & mechanisms. 2009;2(3-4):168-77.

219.Cunningham K, Uchida Y, O'Donnell E, Claudio E, Li W, Soneji K, et al. Conditional deletion of Ccm2 causes hemorrhage in the adult brain: a mouse model

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

of human cerebral cavernous malformations. Human molecular genetics. 2011;20(16):3198-206.

220.Laberge-le Couteulx S, Jung HH, Labauge P, Houtteville JP, Lescoat C, Cecillon M, et al. Truncating mutations in CCM1, encoding KRIT1, cause hereditary cavernous angiomas. Nature genetics. 1999;23(2):189-93.

221.Sahoo T, Johnson EW, Thomas JW, Kuehl PM, Jones TL, Dokken CG, et al. Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (CCM1). Human molecular genetics. 1999;8(12):2325-33.

222.Liquori CL, Berg MJ, Siegel AM, Huang E, Zawistowski JS, Stoffer T, et al. Mutations in a gene encoding a novel protein containing a phosphotyrosine-binding domain cause type 2 cerebral cavernous malformations. American journal of human genetics. 2003;73(6):1459-64.

223.Denier C, Goutagny S, Labauge P, Krivosic V, Arnoult M, Cousin A, et al. Mutations within the MGC4607 gene cause cerebral cavernous malformations. American journal of human genetics. 2004;74(2):326-37.

224.Bergametti F, Denier C, Labauge P, Arnoult M, Boetto S, Clanet M, et al. Mutations within the programmed cell death 10 gene cause cerebral cavernous malformations. American journal of human genetics. 2005;76(1):42-51.

225.Batra S, Lin D, Recinos PF, Zhang J, Rigamonti D. Cavernous malformations: natural history, diagnosis and treatment. Nature reviews Neurology. 2009;5(12):659-70.

226.Dalyai RT, Ghobrial G, Awad I, Tjoumakaris S, Gonzalez LF, Dumont AS, et al. Management of incidental cavernous malformations: a review. Neurosurgical focus. 2011;31(6):E5.

227.Clatterbuck RE, Eberhart CG, Crain BJ, Rigamonti D. Ultrastructural and immunocytochemical evidence that an incompetent blood-brain barrier is related to the pathophysiology of cavernous malformations. Journal of neurology, neurosurgery, and psychiatry. 2001;71(2):188-92.

228.Snellings DA, Hong CC, Ren AA, Lopez-Ramirez MA, Girard R, Srinath A, et al. Cerebral Cavernous Malformation: From Mechanism to Therapy. Circulation research. 2021;129(1):195-215.

229.Phillips CM, Stamatovic SM, Keep RF, Andjelkovic AV. Cerebral Cavernous Malformation Pathogenesis: Investigating Lesion Formation and Progression with Animal Models. International journal of molecular sciences. 2022;23(9).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

230.Bicknell JM, Carlow TJ, Kornfeld M, Stovring J, Turner P. Familial cavernous angiomas. Archives of neurology. 1978;35(11):746-9.

231.Clark JV. Familial occurrence of cavernous angiomata of the brain. Journal of neurology, neurosurgery, and psychiatry. 1970;33(6):871-6.

232.Kidd HA, Cumings JN. Cerebral angiomata in an Icelandic family. Lancet (London, England). 1947;1(6457):747.

233.Gault J, Awad IA, Recksiek P, Shenkar R, Breeze R, Handler M, et al. Cerebral cavernous malformations: somatic mutations in vascular endothelial cells. Neurosurgery. 2009;65(1):138-44; discussion 44-5.

234.Pagenstecher A, Stahl S, Sure U, Felbor U. A two-hit mechanism causes cerebral cavernous malformations: complete inactivation of CCM1, CCM2 or CCM3 in affected endothelial cells. Human molecular genetics. 2009;18(5):911-8.

235.Rath M, Pagenstecher A, Hoischen A, Felbor U. Postzygotic mosaicism in cerebral cavernous malformation. Journal of medical genetics. 2020;57(3):212-6.

236.Malinverno M, Maderna C, Abu Taha A, Corada M, Orsenigo F, Valentino M, et al. Endothelial cell clonal expansion in the development of cerebral cavernous malformations. Nature communications. 2019;10(1):2761.

237.Orsenigo F, Conze LL, Jauhiainen S, Corada M, Lazzaroni F, Malinverno M, et al. Mapping endothelial-cell diversity in cerebral cavernous malformations at single-cell resolution. eLife. 2020;9.

238.Gault J, Shenkar R, Recksiek P, Awad IA. Biallelic somatic and germ line CCM1 truncating mutations in a cerebral cavernous malformation lesion. Stroke. 2005;36(4):872-4.

239.Akers AL, Johnson E, Steinberg GK, Zabramski JM, Marchuk DA. Biallelic somatic and germline mutations in cerebral cavernous malformations (CCMs): evidence for a two-hit mechanism of CCM pathogenesis. Human molecular genetics. 2009;18(5):919-30.

240.McDonald DA, Shi C, Shenkar R, Gallione CJ, Akers AL, Li S, et al. Lesions from patients with sporadic cerebral cavernous malformations harbor somatic mutations in the CCM genes: evidence for a common biochemical pathway for CCM pathogenesis. Human molecular genetics. 2014;23(16):4357-70.

241.Detter MR, Snellings DA, Marchuk DA. Cerebral Cavernous Malformations Develop Through Clonal Expansion of Mutant Endothelial Cells. Circulation research. 2018;123(10):1143-51.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

242.Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013;8(11):2281-308.

243.Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic acids research. 2018;46(W1):W242-w5.

244.Lee DF, Su J, Sevilla A, Gingold J, Schaniel C, Lemischka IR. Combining competition assays with genetic complementation strategies to dissect mouse embryonic stem cell self-renewal and pluripotency. Nature protocols. 2012;7(4):729-48.

245.Katter K, Geurts AM, Hoffmann O, Mátés L, Landa V, Hiripi L, et al. Transposon-mediated transgenesis, transgenic rescue, and tissue-specific gene expression in rodents and rabbits. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2013;27(3):930-41.

246.Sambrook J RD. Molecular cloning: a laboratory manual. 3rd ed. ed New York: Cold Spring Harbor Laboratory Press. 2001.

247.Fuentes-Iglesias A, Ameneiro C, Guallar D, Fidalgo M. Detecting and Modulating ER Stress to Improve Generation of Induced Pluripotent Stem Cells. Methods in molecular biology (Clifton, NJ). 2022;2454:743-54.

248.Covelo-Molares H, Souto Y, Fidalgo M, Guallar D. A Simple, Rapid, and Cost-Effective Method for Loss-of-Function Assays in Pluripotent Cells. Methods in molecular biology (Clifton, NJ). 2022;2520:199-213.

249.Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature. 2013;495(7441):370-4.

250.Štefková K, Procházková J, Pacherník J. Alkaline phosphatase in stem cells. Stem cells international. 2015;2015:628368.

251.Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. Journal of embryology and experimental morphology. 1985;87:27-45.

252.Beccari L, Moris N, Girgin M, Turner DA, Baillie-Johnson P, Cossy AC, et al. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature. 2018;562(7726):272-6.

253.Baillie-Johnson P, van den Brink SC, Balayo T, Turner DA, Martinez Arias A. Generation of Aggregates of Mouse Embryonic Stem Cells that Show Symmetry

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Breaking, Polarization and Emergent Collective Behaviour In Vitro. Journal of visualized experiments : JoVE. 2015(105).

254.Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, et al. Extracting biological meaning from large gene lists with DAVID. Current protocols in bioinformatics. 2009;Chapter 13:Unit 13.1.

255.Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic acids research. 2019;47(W1):W212-w24.

256.Wang Z, Lachmann A, Keenan AB, Ma'ayan A. L1000FWD: fireworks visualization of drug-induced transcriptomic signatures. Bioinformatics (Oxford, England). 2018;34(12):2150-2.

257.Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nature protocols. 2006;1(2):729-48.

258.Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nature structural & molecular biology. 2016;23(4):349-57.

259.Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.

260.Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature protocols. 2009;4(8):1184-91.

261.Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

262.Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576-89.

263.Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics (Oxford, England). 2022.

264.Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology. 2008;9(9):R137.

265.Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078-9.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

266.Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15-21.

267.Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research. 2021;49(D1):D605-d12.

268.Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114-20.

269.Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome biology. 2009;10(11):R130.

270.Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, et al. Deterministic and stochastic allele specific gene expression in single mouse blastomeres. PloS one. 2011;6(6):e21208.

271.Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (New York, NY). 2012;337(6096):816-21.

272.Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY). 2013;339(6121):819-23.

273.Orford KW, Scadden DT. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nature reviews Genetics. 2008;9(2):115-28.

274.Boward B, Wu T, Dalton S. Concise Review: Control of Cell Fate Through Cell Cycle and Pluripotency Networks. Stem cells (Dayton, Ohio). 2016;34(6):1427-36.

275.Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, et al. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Developmental cell. 2015;35(3):366-82.

276.Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-19.

277.Keshet Y, Seger R. The MAP kinase signaling cascades: a system of hundreds of components regulates a diverse array of physiological functions. Methods in molecular biology (Clifton, NJ). 2010;661:3-38.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

278.Kunath T, Saba-El-Leil MK, Almousailleakh M, Wray J, Meloche S, Smith A. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development (Cambridge, England). 2007;134(16):2895-902.

279.Xu P, Davis RJ. c-Jun NH2-terminal kinase is required for lineage-specific differentiation but not stem cell self-renewal. Molecular and cellular biology. 2010;30(6):1329-40.

280.Williams CA, Fernandez-Alonso R, Wang J, Toth R, Gray NS, Findlay GM. Erk5 Is a Key Regulator of Naive-Primed Transition and Embryonic Stem Cell Identity. Cell reports. 2016;16(7):1820-8.

281.Liu Y, Yamane J, Tanaka A, Fujibuchi W, Yamashita JK. AMPK activation reverts mouse epiblast stem cells to naive state. iScience. 2021;24(7):102783.

282.Lavoie H, Gagnon J, Therrien M. ERK signalling: a master regulator of cell behaviour, life and fate. Nature reviews Molecular cell biology. 2020;21(10):607-32.

283.Choi J, Huebner AJ, Clement K, Walsh RM, Savol A, Lin K, et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature. 2017;548(7666):219-23.

284.Hackett JA, Kobayashi T, Dietmann S, Surani MA. Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. Stem cell reports. 2017;8(6):1645-58.

285.Roskoski R, Jr. ERK1/2 MAP kinases: structure, function, and regulation. Pharmacological research. 2012;66(2):105-43.

286.Loh CH, Veenstra GJC. The Role of Polycomb Proteins in Cell Lineage Commitment and Embryonic Development. Epigenomes. 2022;6(3).

287.Das PP, Hendrix DA, Apostolou E, Buchner AH, Canver MC, Beyaz S, et al. PRC2 Is Required to Maintain Expression of the Maternal Gtl2-Rian-Mirg Locus by Preventing De Novo DNA Methylation in Mouse Embryonic Stem Cells. Cell reports. 2015;12(9):1456-70.

288.Clark AT. DNA methylation remodeling in vitro and in vivo. Current opinion in genetics & development. 2015;34:82-7.

289.Zhang Y, Xiang Y, Yin Q, Du Z, Peng X, Wang Q, et al. Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nature genetics. 2018;50(1):96-105.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

290.Lauria A, Meng G, Proserpio V, Rapelli S, Maldotti M, Polignano IL, et al. DNMT3B supports meso-endoderm differentiation from mouse embryonic stem cells. Nature communications. 2023;14(1):367.

291.Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015;520(7546):243-7.

292.Arab K, Karaulanov E, Musheev M, Trnka P, Schäfer A, Grummt I, et al. GADD45A binds R-loops and recruits TET1 to CpG island promoters. Nature genetics. 2019;51(2):217-23.

293.Rasmussen KD, Berest I, Keßler S, Nishimura K, Simón-Carrasco L, Vassiliou GS, et al. TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. Genome research. 2019;29(4):564-75.

294.Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA methylation prevents spurious transcription initiation. Nature. 2017;543(7643):72-7.

295.Reimer M, Jr., Pulakanti K, Shi L, Abel A, Liang M, Malarkannan S, et al. Deletion of Tet proteins results in quantitative disparities during ESC differentiation partially attributable to alterations in gene expression. BMC developmental biology. 2019;19(1):16.

296.Boroviak T, Loos R, Bertone P, Smith A, Nichols J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nature cell biology. 2014;16(6):516-28.

297.Li M, Yu JSL, Tilgner K, Ong SH, Koike-Yusa H, Yusa K. Genome-wide CRISPR-KO Screen Uncovers mTORC1-Mediated Gsk3 Regulation in Naive Pluripotency Maintenance and Dissolution. Cell reports. 2018;24(2):489-502.

298.Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, et al. Dissecting self-renewal in stem cells with RNA interference. Nature. 2006;442(7102):533-8.

299.Hailesellasse Sene K, Porter CJ, Palidwor G, Perez-Iratxeta C, Muro EM, Campbell PA, et al. Gene function in early mouse embryonic stem cell differentiation. BMC genomics. 2007;8:85.

300.Blancas AA, Lauer NE, McCloskey KE. Endothelial differentiation of embryonic stem cells. Current protocols in stem cell biology. 2008;Chapter 1:Unit 1F.5.

301.Blancas AA, Shih AJ, Lauer NE, McCloskey KE. Endothelial cells from embryonic stem cells in a chemically defined medium. Stem cells and development. 2011;20(12):2153-61.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

302.Chiang PM, Wong PC. Differentiation of an embryonic stem cell to hemogenic endothelium by defined factors: essential role of bone morphogenetic protein 4. Development (Cambridge, England). 2011;138(13):2833-43.

303.Furuta K, Guo Q, Pavelko KD, Lee JH, Robertson KD, Nakao Y, et al. Lipid-induced endothelial vascular cell adhesion molecule 1 promotes nonalcoholic steatohepatitis pathogenesis. The Journal of clinical investigation. 2021;131(6).

304.Harland LTG, Simon CS, Senft AD, Costello I, Greder L, Imaz-Rosshandler I, et al. The T-box transcription factor Eomesodermin governs haemogenic competence of yolk sac mesodermal progenitors. Nature cell biology. 2021;23(1):61-74.

305.Marchetti L, Francisco D, Soldati S, Haghayegh Jahromi N, Barcos S, Gruber I, et al. ACKR1 favors transcellular over paracellular T-cell diapedesis across the blood-brain barrier in neuroinflammation in vitro. European journal of immunology. 2022;52(1):161-77.

306.Rahman M, Ding Z, Rönnow CF, Thorlacius H. Transcriptomic Analysis Reveals Differential Expression of Genes between Lung Capillary and Post Capillary Venules in Abdominal Sepsis. International journal of molecular sciences. 2021;22(19).

307.Das A, Fernandez-Zapico ME, Cao S, Yao J, Fiorucci S, Hebbel RP, et al. Disruption of an SP2/KLF6 repression complex by SHP is required for farnesoid X receptor-induced endothelial cell migration. The Journal of biological chemistry. 2006;281(51):39105-13.

308.Lee JH, Chun T, Park SY, Rho SB. Interferon regulatory factor-1 (IRF-1) regulates VEGF-induced angiogenesis in HUVECs. Biochimica et biophysica acta. 2008;1783(9):1654-62.

309.Lan Y, Liu B, Yao H, Li F, Weng T, Yang G, et al. Essential role of endothelial Smad4 in vascular remodeling and integrity. Molecular and cellular biology. 2007;27(21):7683-92.

310.Li F, Lan Y, Wang Y, Wang J, Yang G, Meng F, et al. Endothelial Smad4 maintains cerebrovascular integrity by activating N-cadherin through cooperation with Notch. Developmental cell. 2011;20(3):291-302.

311.Mammoto A, Muyleart M, Kadlec A, Gutterman D, Mammoto T. YAP1-TEAD1 signaling controls angiogenesis and mitochondrial biogenesis through PGC1α. Microvascular research. 2018;119:73-83.

312.Yang J, Zhang L, Jiang Z, Ge C, Zhao F, Jiang J, et al. TCF12 promotes the tumorigenesis and metastasis of hepatocellular carcinoma via upregulation of CXCR4 expression. Theranostics. 2019;9(20):5810-27.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

313.Fan CS, Chen WS, Chen LL, Chen CC, Hsu YT, Chua KV, et al. Osteopontin-integrin engagement induces HIF-1α-TCF12-mediated endothelial-mesenchymal transition to exacerbate colorectal cancer. Oncotarget. 2018;9(4):4998-5015.

314.Hao JS, Zhu CJ, Yan BY, Yan CY, Ling R. Stimulation of KLF14/PLK1 pathway by thrombin signaling potentiates endothelial dysfunction in Type 2 diabetes mellitus. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2018;99:859-66.

315.Duan HX, Li BW, Zhuang X, Wang LT, Cao Q, Tan LH, et al. TCF21 inhibits tumor-associated angiogenesis and suppresses the growth of cholangiocarcinoma by targeting PI3K/Akt and ERK signaling. American journal of physiology Gastrointestinal and liver physiology. 2019;316(6):G763-g73.

316.Tanaka A, Itoh F, Nishiyama K, Takezawa T, Kurihara H, Itoh S, et al. Inhibition of endothelial cell activation by bHLH protein E2-2 and its impairment of angiogenesis. Blood. 2010;115(20):4138-47.

317.Meng ZZ, Liu W, Xia Y, Yin HM, Zhang CY, Su D, et al. The pro-inflammatory signalling regulator Stat4 promotes vasculogenesis of great vessels derived from endothelial precursors. Nature communications. 2017;8:14640.

318.Tuli HS, Sak K, Iqubal A, Garg VK, Varol M, Sharma U, et al. STAT signaling as a target for intervention: from cancer inflammation and angiogenesis to non-coding RNAs modulation. Molecular biology reports. 2022;49(9):8987-99.

319.Akhter MS, Uddin MA, Barabutis N. P53 Regulates the Redox Status of Lung Endothelial Cells. Inflammation. 2020;43(2):686-91.

320.Appukuttan B, McFarland TJ, Stempel A, Kassem JB, Hartzell M, Zhang Y, et al. The related transcriptional enhancer factor-1 isoform, TEAD4(216), can repress vascular endothelial growth factor expression in mammalian cells. PloS one. 2012;7(6):e31260.

321.Ding X, Yao W, Zhu J, Mu K, Zhang J, Zhang JA. Resveratrol Attenuates High Glucose-Induced Vascular Endothelial Cell Injury by Activating the E2F3 Pathway. BioMed research international. 2020;2020:6173618.

322.Zhou J, Cheng M, Wu M, Boriboun C, Jujo K, Xu S, et al. Contrasting roles of E2F2 and E2F3 in endothelial cell growth and ischemic angiogenesis. Journal of molecular and cellular cardiology. 2013;60:68-71.

323.Burnstock G. Endothelium-derived vasoconstriction by purines and pyrimidines. Circulation research. 2008;103(10):1056-7.

324.Ding A, Bian YY, Zhang ZH. SP1/TGF-β1/SMAD2 pathway is involved in angiogenesis during osteogenesis. Molecular medicine reports. 2020;21(3):1581-9.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

325.Itoh F, Itoh S, Adachi T, Ichikawa K, Matsumura Y, Takagi T, et al. Smad2/Smad3 in endothelium is indispensable for vascular stability via S1PR1 and N-cadherin expressions. Blood. 2012;119(22):5320-8.

326.Smits M, Wurdinger T, van het Hof B, Drexhage JA, Geerts D, Wesseling P, et al. Myc-associated zinc finger protein (MAZ) is regulated by miR-125b and mediates VEGF-induced angiogenesis in glioblastoma. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2012;26(6):2639-47.

327.Senoo M, Matsumura Y, Habu S. TAp63gamma (p51A) and dNp63alpha (p73L), two major isoforms of the p63 gene, exert opposite effects on the vascular endothelial growth factor (VEGF) gene expression. Oncogene. 2002;21(16):2455-65.

328.Sun X, Hu F, Hou Z, Chen Q, Lan J, Luo X, et al. SIX4 activates Akt and promotes tumor angiogenesis. Experimental cell research. 2019;383(1):111495.

329.Hübner K, Grassme KS, Rao J, Wenke NK, Zimmer CL, Korte L, et al. Wnt signaling positively regulates endothelial cell fate specification in the Fli1a-positive progenitor population via Lef1. Developmental biology. 2017;430(1):142-55.

330.Mahamud MR, Geng X, Ho YC, Cha B, Kim Y, Ma J, et al. GATA2 controls lymphatic endothelial cell junctional integrity and lymphovenous valve morphogenesis through miR-126. Development (Cambridge, England). 2019;146(21).

331.Cutsforth-Gregory JK, Lanzino G, Link MJ, Brown RD, Jr., Flemming KD. Characterization of radiation-induced cavernous malformations and comparison with a nonradiation cavernous malformation cohort. Journal of neurosurgery. 2015;122(5):1214-22.

332.Koike T, Yanagimachi N, Ishiguro H, Yabe H, Yabe M, Morimoto T, et al. High incidence of radiation-induced cavernous hemangioma in long-term survivors who underwent hematopoietic stem cell transplantation with radiation therapy during childhood or adolescence. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2012;18(7):1090-8.

333.Vinchon M, Leblond P, Caron S, Delestret I, Baroncini M, Coche B. Radiation-induced tumors in children irradiated for brain tumor: a longitudinal study. Child's nervous system : ChNS : official journal of the International Society for Pediatric Neurosurgery. 2011;27(3):445-53.

334.Polster SP, Sharma A, Tanes C, Tang AT, Mericko P, Cao Y, et al. Permissive microbiome characterizes human subjects with a neurovascular disease cavernous angioma. Nature communications. 2020;11(1):2659.

335.Tang AT, Choi JP, Kotzin JJ, Yang Y, Hong CC, Hobson N, et al. Endothelial TLR4 and the microbiome drive cerebral cavernous malformations. Nature. 2017;545(7654):305-10.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

336.Qiu J, Hirschi KK. Endothelial Cell Development and Its Application to Regenerative Medicine. Circulation research. 2019;125(4):489-501.

337.Abhinand CS, Raju R, Soumya SJ, Arya PS, Sudhakaran PR. VEGF-A/VEGFR2 signaling network in endothelial cells relevant to angiogenesis. Journal of cell communication and signaling. 2016;10(4):347-54.

338.Chen G, Yin S, Zeng H, Li H, Wan X. Regulation of Embryonic Stem Cell Self-Renewal. Life (Basel, Switzerland). 2022;12(8).

339.Chen G, Xu X, Zhang L, Fu Y, Wang M, Gu H, et al. Blocking autocrine VEGF signaling by sunitinib, an anti-cancer drug, promotes embryonic stem cell self-renewal and somatic cell reprogramming. Cell research. 2014;24(9):1121-36.

340.Lee SW, Jeong HK, Lee JY, Yang J, Lee EJ, Kim SY, et al. Hypoxic priming of mESCs accelerates vascular-lineage differentiation through HIF1-mediated inverse regulation of Oct4 and VEGF. EMBO molecular medicine. 2012;4(9):924-38.

341.Yoshida Y, Takahashi K, Okita K, Ichisaka T, Yamanaka S. Hypoxia enhances the generation of induced pluripotent stem cells. Cell stem cell. 2009;5(3):237-41.

342.Garcia MD, Larina IV. Vascular development and hemodynamic force in the mouse yolk sac. Frontiers in physiology. 2014;5:308.

343.Watari K, Shibata T, Fujita H, Shinoda A, Murakami Y, Abe H, et al. NDRG1 activates VEGF-A-induced angiogenesis through PLCγ1/ERK signaling in mouse vascular endothelial cells. Communications biology. 2020;3(1):107.

344.Fish JE, Cantu Gutierrez M, Dang LT, Khyzha N, Chen Z, Veitch S, et al. Dynamic regulation of VEGF-inducible genes by an ERK/ERG/p300 transcriptional network. Development (Cambridge, England). 2017;144(13):2428-44.

345.Chen P, Zhang H, Zhang Q, Zhou W, Deng Y, Hu X, et al. Basic Fibroblast Growth Factor Reduces Permeability and Apoptosis of Human Brain Microvascular Endothelial Cells in Response to Oxygen and Glucose Deprivation Followed by Reoxygenation via the Fibroblast Growth Factor Receptor 1 (FGFR1)/ERK Pathway. Medical science monitor : international medical journal of experimental and clinical research. 2019;25:7191-201.

346.Zhu H, Jin Q, Li Y, Ma Q, Wang J, Li D, et al. Melatonin protected cardiac microvascular endothelial cells against oxidative stress injury via suppression of IP3R-[Ca(2+)]c/VDAC-[Ca(2+)]m axis by activation of MAPK/ERK signaling pathway. Cell stress & chaperones. 2018;23(1):101-13.

347.Charron J, Bissonauth V, Nadeau V. Implication of MEK1 and MEK2 in the establishment of the blood-placenta barrier during placentogenesis in mouse. Reproductive biomedicine online. 2012;25(1):58-67.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

348.Bissonauth V, Roy S, Gravel M, Guillemette S, Charron J. Requirement for Map2k1 (Mek1) in extra-embryonic ectoderm during placentogenesis. Development (Cambridge, England). 2006;133(17):3429-40.

349.Hatano N, Mori Y, Oh-hora M, Kosugi A, Fujikawa T, Nakai N, et al. Essential role for ERK2 mitogen-activated protein kinase in placental development. Genes to cells : devoted to molecular & cellular mechanisms. 2003;8(11):847-56.

350.Yao Y, Li W, Wu J, Germann UA, Su MS, Kuida K, et al. Extracellular signal-regulated kinase 2 is necessary for mesoderm differentiation. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(22):12759-64.

351.Giroux S, Tremblay M, Bernard D, Cardin-Girard JF, Aubry S, Larouche L, et al. Embryonic death of Mek1-deficient mice reveals a role for this kinase in angiogenesis in the labyrinthine region of the placenta. Current biology : CB. 1999;9(7):369-72.

352.Oki S, Kitajima K, Meno C. Dissecting the role of Fgf signaling during gastrulation and left-right axis formation in mouse embryos using chemical inhibitors. Developmental dynamics : an official publication of the American Association of Anatomists. 2010;239(6):1768-78.

353.Hacohen Lev-Ran A, Seger R. Retention of ERK in the cytoplasm mediates the pluripotency of embryonic stem cells. Stem cell reports. 2023;18(1):305-18.

354.Richter GH, Plehm S, Fasan A, Rössler S, Unland R, Bennani-Baiti IM, et al. EZH2 is a mediator of EWS/FLI1 driven tumor growth and metastasis blocking endothelial and neuro-ectodermal differentiation. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(13):5324-9.

355.Huang X, Bashkenova N, Hong Y, Lyu C, Guallar D, Hu Z, et al. A TET1-PSPC1-Neat1 molecular axis modulates PRC2 functions in controlling stem cell bivalency. Cell reports. 2022;39(10):110928.

356.Wattanathamsan O, Chetprayoon P, Chantaravisoot N, Wongkongkathep P, Chanvorachote P, Pongrakhananon V. CAMSAP3 depletion induces lung cancer cell senescence-associated phenotypes through extracellular signal-regulated kinase inactivation. Cancer medicine. 2021;10(24):8961-75.

357.Kim SY, Choi J, Lee DH, Park JH, Hwang YJ, Baek KH. PME-1 is regulated by USP36 in ERK and Akt signaling pathways. FEBS letters. 2018;592(9):1575-88.

358.Enomoto A, Kido N, Ito M, Morita A, Matsumoto Y, Takamatsu N, et al. Negative regulation of MEKK1/2 signaling by serine-threonine kinase 38 (STK38). Oncogene. 2008;27(13):1930-8.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

359.Jin C, Chen Z, Shi W, Lian Q. Tropomodulin 3 promotes liver cancer progression by activating the MAPK/ERK signaling pathway. Oncology reports. 2019;41(5):3060-8.

360.García Ponce A, Citalán Madrid AF, Vargas Robles H, Chánez Paredes S, Nava P, Betanzos A, et al. Loss of cortactin causes endothelial barrier dysfunction via disturbed adrenomedullin secretion and actomyosin contractility. Scientific reports. 2016;6:29003.

361.Valencia-Expósito A, Grosheva I, Míguez DG, González-Reyes A, Martín-Bermudo MD. Myosin light-chain phosphatase regulates basal actomyosin oscillations during morphogenesis. Nature communications. 2016;7:10746.

362.Angulo-Urarte A, Casado P, Castillo SD, Kobialka P, Kotini MP, Figueiredo AM, et al. Endothelial cell rearrangements during vascular patterning require PI3-kinase-mediated inhibition of actomyosin contractility. Nature communications. 2018;9(1):4826.

363.Boraas LC, Pineda ET, Ahsan T. Actin and myosin II modulate differentiation of pluripotent stem cells. PloS one. 2018;13(4):e0195588.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Las proteínas CCM son esenciales para un correcto desarrollo embrionario y función endotelial, pero todavía se desconoce el papel que desempeñan en pluripotencia y en la toma de decisiones celulares que determinan el linaje endotelial. Mis resultados revelan que las proteínas CCM1, CCM2 y CCM3 intervienen en el control transcripcional que gobierna pluripotencia y que son necesarias durante la diferenciación celular, incluyendo la especificación endotelial a partir de ESCs. Además, muestro que CCM2 coordina estados alternativos de pluripotencia a través de la vía de señalización de ERK1/2 y de programas epigenéticos. Los mecanismos identificados en esta tesis doctoral abren nuevas vías de estudio para entender mejor la función CCM en desarrollo temprano, biología endotelial y enfermedad.